The development of aptamer-based probes for the detection of TB antigens ESAT-6.CFP-10: potential TB diagnostic tools

By

MATSOPIANE CHARLOTTE MASERUMULE

(Student number: MSRMAT002)

Dissertation presented for the degree of Master of Science

MSc Medicine (MD1)

In the Department of Medicine

UNIVERSITY OF CAPE TOWN

14th August 2013

Supervisors:

Prof. Keertan Dheda

Dr. Lionel Gresh

Dr. Makobetsa Khati

DECLARATION

I, MATSOPIANE CHARLOTTE MASERUMULE, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate……

Date: 14th August 2013

ABSTRACT

Tuberculosis (TB) is one of the most life-threatening infectious diseases in the world, claiming millions of lives annually. Lack of point-of-care (PoC) diagnostic tools for TB hinders control of the disease, particularly in resource-limited, high HIV and TB prevalence countries. Therefore, there is a need for simple, rapid, accurate, and affordable PoC diagnostics to detect active TB early enough for opportune intervention. To develop TB detection probes that will constitute such diagnostics, our research group recently isolated DNA aptamers that bind to a putative marker for active TB; the ESAT-6.CFP-10 heterodimer. Aptamers are highly specific artificial mimics of antibodies that have shown great prospects in diagnostic applications. The aim of this study was to characterise the anti-ESAT-6.CFP-10 aptamers, and to optimise them into more specific and affordable detection probes for the development of potential PoC TB diagnostic tools. Characterisation included determining the binding kinetics of six full length (90-mer) anti-ESAT-6.CFP-10 aptamers by Surface Plasmon Resonance (SPR) technology, and the analysis of the secondary structures and associated energetics using *in silico* methods. Five of the full length anti-ESAT-6.CFP-10 aptamers exhibited high affinity to the recombinant CFP-10 monomer, with dissociation constant (K_D) values within the nanomolar range. *In silico* secondary structure analyses provided a theoretical framework for the target-binding motifs of the aptamer sequences. Optimisation of two of these aptamers, namely CSIR 2.11 and CSIR 2.19, was carried out by the rational truncation of the full length sequences, retaining their predicted minimum target-binding motifs. Rational truncation yielded two shorter derivatives of each aptamer; 70-mer and 77-mer for CSIR 2.11, and 77-and 49-mer for CSIR 2.19. *In silico* conformational energy landscape analyses suggested that the truncated aptamers had improved folding kinetics compared to the full length parent aptamers. When evaluated in an enzyme linked oligonucleotide assay (ELONA), the functional performance of the truncated aptamers was comparable to that of the full length aptamers. These data suggest the potential utility of the truncated aptamers as candidate TB detection probes for the design of a PoC TB diagnostic tool.

DEDICATION

This dissertation is dedicated to the memory of my late parents; my mother Angeline Nai Maserumule, and my father Ernest Mogotwane Maserumule. They both held education and independence of thought in high regard. Their philosophy has fuelled my all of my endeavours.

RESEARCH OUTPUTS

Publications

- Rotherham LS, **Maserumule C**, Dheda K, Theron J, Khati M (**2012**) Selection and Application of ssDNA Aptamers to Detect Active TB from Sputum Samples. PLoS ONE 7(10): e46862. doi:10.1371/journal.pone.0046862

Patents

- Rotherham, L, **Maserumule, C** and M. Khati **(2011).** High affinity nucleic acid aptamer ligands − useful for diagnosis of TB South Africa P53037ZP00 (provisional patent: converted).

- Rotherham, L., **Maserumule, C.**, and Khati, M. (**2012**). High affinity nucleic acid aptamer ligands − useful for diagnosis of TB. PCT/IB2012/055025 (PCT application).

Presentations at conferences and workshops

- **Matsopiane Charlotte Maserumule**, Lia Rotherham, Lionel Gresh, Keertan Dheda, Makobetsa Khati. The development of aptamer-based probes for the detection of TB antigens ESAT-6/CFP-10: potential TB diagnostic tools. The University of Cape Town's 39th Annual Medicine Research Day, October **2012**, Cape Town, South Africa (*Oral*, *Won the Bernard Pimstone Best Presentation Award for Laboratory Medicine*).

- **Matsopiane Charlotte Maserumule,** The KwaZulu-Natal Research Institute for TB and HIV (K-RIT) Official opening and Scientific Symposium, October **2012**, Durban, South Africa (*Oral*).

- **Matsopiane Charlotte Maserumule**, Lia Rotherham, Lionel Gresh, Keertan Dheda, Makobetsa Khati. The development of aptamer-based probes for the detection of TB antigens ESAT-6/CFP-10: potential TB diagnostic tools. The South African TB conference, June **2012** (ICC, Durban, South Africa (*Poster*).

- **Matsopiane Charlotte Maserumule.** The ACGT Second Regional Synthetic Biology Forum, July **2012**, Johannesburg, South Africa (*Poster*).
- Lia Rotherham, **Charlotte Maserumule**, Keertan Dheda, Jacques Theron and Makobetsa Khati. Detection of active TB in clinical sputum samples using ssDNA aptamers: a potential Point-of-Care diagnostic. CSIR Biennial Conference, October **2012**, CSIR Pretoria, South Africa (*Poster*).
- Lia Rotherham, Lionel Gresh, **Charlotte Maserumule**, Keertan Dheda, Jacques Theron and Makobetsa Khati. Isolation and Characterisation of Novel Aptamers against an Antigen of Mycobacterium Tuberculosis for the Development of a Rapid and Accurate TB Diagnostic tool. Keystone meeting; Tuberculosis: Immunology, Cell Biology and Novel Vaccination Strategies, 15 – 20 January **2011**, Vancouver, British Columbia Canada (*Poster*).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AFB	acid-fast bacilli
AuNP	gold nanoparticles
BCA	bicinchoninic acid
BCG	Bacillus Calmette-Guérin
bp	base pair
CFP-10	culture filtrate protein-10
CFP-10/ESAT-6	culture filtrate protein-10 and early secreted antigenic target, 6 kDa heterodimer
DNA	deoxyribonucleic acid
dsDNA	double-stranded deoxyribonucleic acid
DTT	dithiothreitol
ECL	enhanced chemiluminescence
E. coli	*Escherichia coli*
EDC	N-ethyl-N-(3,-dimethylaminopropyl)-carbodiimide hydrochloride
ELISA	enzyme-linked immunosorbent assay
ELONA	enzyme-linked oligonucleotide assay
ESAT-6	early secreted antigenic target, 6 kDa
HBS-N	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, sodium chloride buffer
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	human immunodeficiency virus
IgG	immunoglobulin G
IGRA	interferon-γ-release assay
IPTG	isopropyl-1-thio-β-D-glactopyranoside
K_D	dissociation constant
kDa	kilo Dalton
KOH	potassium hydroxide

LB	Luria-Bertani broth
M. bovis	*Mycobacterium bovis*
M. smeg.	*Mycobacterium smegmatis*
MTB	*Mycobacterium tuberculosis*
MFE	minimum free energy
NAATs	nucleic acid amplification tests
Ni-NTA	nickel nitrilotriacetic acid
NHS	N-hydroxysuccinimide
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PBS-T	phosphate-buffered saline containing 0.005% Tween-20
PCR	polymerase chain reaction
PSMA	prostate specific membrane antigen
PoC	point-of-care
RD	region of difference
RNA	ribonucleic acid
rpm	revolutions per minute
RU	response units
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SELEX	systematic evolution of ligands exponential enrichment
SPR	surface plasmon resonance
ssDNA	single-stranded deoxyribonucleic acid
TB	Tuberculosis
TMB	tetramethylbenzidine
TST	tuberculin skin test
VEGF	vascular endothelial growth factor
WHO	World Health Organisation

ACKNOWLEDGMENTS

My deepest gratitude goes to my supervisors Dr. Lionel Gresh, Dr. Makobetsa Khati, and Prof. Keertan Dheda for their guidance, support and encouragement throughout my MSc studies. Special thanks to Dr Khati for affording me the opportunity to be part of the Aptamer Team. The opportunity provided the basis for my further development as a Scientist and was a life changing experience. I am thankful to Prof Dheda for his encouraging feedback and direction. I am also immensely grateful to Dr Lionel Gresh, for his invaluable mentorship. Dr Gresh created a nurturing environment, which allowed me the space to learn, explore, and find my feet. He has taught me a great deal more without overtly spelling out the lessons. I have learned from him, experientially, the art of being more meticulous in research; starting with the basics, to asking very specific questions, and not getting lost in the quest to find answers. He also modelled the flair of shrugging off the inevitable mishaps and embracing the surprises along the way and staying on course. I appreciate my colleagues in the Aptamer Technology Research Group for their support and guidance in my research and related endeavours, and for easing the tough times with bouts of laughter and words of encouragement. Special thanks to Dr Lia Rotherham whose PhD work yielded the original aptamer sequences used in my master's research; for sharing her resources and knowhow. Heartfelt thanks to my family and friends for their reassuring support. I am indebted to my older sister, Portia Maserumule and to my niece and dearest friend, Dolly Makola; for their immeasurable support. Many thanks to the Department of Medicine, UCT and Groote Schuur Hospital, for hosting me as a student; Biosciences Unit, CSIR; the DST, South Africa; and the K-RITH SA Collaborative Grants programme student support grant for funding my research and related activities. Finally, all thanks be to the One who wove it all together, the Lord God Almighty; for the blessing of this opportunity, for being the wind beneath my wings and my pillar of strength, for opening exciting avenues (both sought and unsought) on this dynamic adventure, and for keeping me sane through the madness of Life Science research.

CHAPTER 1

INTRODUCTION

1.1 The Epidemiology of Tuberculosis

Tuberculosis (TB) remains a major global health burden associated with high rates of morbidity and mortality. There are approximately 8.7 million new cases of clinically active TB, and about 1.4 million TB-attributed deaths annually (WHO, 2012). Of the 8.7 million new active TB cases, 13% were coinfected with human immunodeficiency virus (HIV) (WHO, 2012) . Furthermore, of the reported global 1.4 million deaths in 2011, almost a million were among HIV-negative individuals and 430 000 among those infected HIV (WHO, 2012). Globally, TB ranks as the second leading infectious disease (after HIV) that causes death; with the gravest burden experienced by Asia and Africa (WHO, 2012). This is despite the disease being largely curable and the progressive improvements in access to TB treatment in the last decade.

1.2 The Pathogenesis of TB

TB is an air-borne infectious disease caused by a group of closely related species which collectively form the *Mycobacterium tuberculosis* complex. These species include *M. tuberculosis, M. bovis, M. africanum, M. microti, and M. canettii* (Fukuda et al., 2013). *M. tuberculosis* (MTB) is the primary causative agent for the majority of human TB cases worldwide (Fukuda et al., 2013). This aerobic, slow-growing, rod-shaped, acid-fast bacterium (Dinic et al., 2013) has a complex cell wall structure; which is known to play a role in both its pathogenesis and resistance to treatment (Knechel, 2009). To elicit pathogenesis, MTB enters the airways of exposed individuals in aerosol droplets generated through coughing, sneezing and talking (Knechel, 2009)The introduction of MTB bacilli into the lungs can cause respiratory infection called pulmonary TB, which can later disseminate to other organs resulting in extrapulmonary TB (Solovic et al., 2013).

Generally, a small proportion of people infected with MTB will progress to develop active TB disease (Smith, 2003). Most healthy individuals successfully control the infection through host immune response mechanisms involving the phagocytotic arrest of the MTB bacilli in granulomas (Smith, 2003). Successful arrest of the infection results in the asymptomatic and non-transmissible state of the disease, termed latent infection (Smith, 2003). However, if the exposed individual has compromised immunity as would be the case in HIV-infected persons, or immature immunity such as in children younger than 5 years, the infection may progress to active TB disease (Knechel, 2009, Smith, 2003). Furthermore, reactivation of the disease , termed secondary progressive TB, may occur in immunocompetent individuals within the first 2 years of initial infection or even decades later, under immunity-stressful conditions (Fukuda et al., 2013).

1.3 The problem of TB diagnosis

Central to the burden of TB is poor diagnostics, particularly in high prevalence countries (Chamie et al., Dorman, 2010, Mathebula et al., 2009, Storla et al., 2008). Delayed or poor diagnosis may lead to continued transmission of the disease in communities and hospitals (Chamie et al., Dorman, 2010). In addition, inadequacies in TB diagnostic tools may lead to false positives, resulting in unnecessary utilization of public health resources and possibly dire health consequences to patients (Chamie et al., Dorman, 2010).

The method most commonly used to diagnose active TB is sputum smear microscopy (Chamie et al., Dorman, 2010, Mayer and Dukes Hamilton, 2010, Davis et al., 2011) The major disadvantages of this method of detection are the requirement of laboratory equipment and trained personnel, as well as poor sensitivity of detection which is estimated at 70% but also as low as 35 % in nations with a high prevalence of HIV-TB coinfection (Davis et al., 2011). Some reports indicate that up to 50% of HIV-positive TB-infected patients show negative smear test results (Parida and Kaufmann, 2010). The

tuberculin skin test (TST) and chest radiographs are usually used in conjunction with smear microscopy (Chamie et al., Dorman, 2010). However, the TST presents low sensitivity for active TB while chest radiography is limited to detection of pulmonary TB, has low sensitivity, and is generally subjective (Dorman, 2010). The diagnosis of extrapulmonary TB often requires invasive measures of obtaining diagnostic samples from the site of infection (Nicol and Zar, 2011, Parida and Kaufmann, 2010) and fairly technical methods of detection (Parida and Kaufmann, 2010). The culturing of MTB, both on solid and in liquid media, is considered the current gold standard in TB diagnosis and is significantly more sensitive than smear microscopy (Zeka et al., 2011). However, this method is time-consuming (Zeka et al., 2011), costly and also requires trained personnel as well as specialized Bio-safety level-3 facilities (Chamie et al., Dorman, 2010, Parida and Kaufmann, 2010).

Newer technologies in TB diagnosis include the interferon γ-release assays (IGRAs) and nucleic acid amplification tests (NAATs). The IGRAs are serodiagnostic assays that offer rapid TB diagnosis based on MTB antigen-induced immune response (de Moraes Van-Lume, 2010, Ernst et al., 2007, Pai and Ling, 2008, Pai et al., 2008). These tests offer a single patient visit (de Moraes Van-Lume, 2010, Ernst et al., 2007) and modestly high sensitivity (de Moraes Van-Lume, 2010, Dheda et al., 2009, Ernst et al., 2007). However, the IGRAs have shown poor specificity as other mycobacteria (e.g. *M. kansasii* and *M. marinum*) can also elicit an interferon γ response, and thus a give a positive IGRA test (Dheda et al., 2009). One of the key shortcomings of these tests is the inability to distinguish active from latent TB, particularly in high burden countries (de Moraes Van-Lume, 2010, Dheda et al., 2009, Ernst et al., 2007). Other limitations of the IGRAs include the requirement of equipment and trained personnel (de Moraes Van-Lume, 2010, Dheda et al., 2009, Ernst et al., 2007).

INTRODUCTION

NAATs use various molecular approaches to amplify and detect the Mycobacterium tuberculosis complex (MTC) directly in clinical specimens, allowing for rapid and accurate TB detection (Antonenka et al., 2013). Commercially available NAATs used in TB diagnosis include the ProbeTec ET DTB (DTB) (BD), DNA based probe Accuprobe (Genprobe), the Amplified M. tuberculosis Direct (AMPLIFIED MTD) test (GenProbe), Amplicor MTB (Roche), COBAS TaqMan MTB (CTM-MTB) (Roche), and the Xpert MTB/ RIF™ (Cepheid) (Boehme et al., 2010, Davis et al., 2011) . Although these tests offer promising potential in TB diagnosis with high specificity, ranging from 85–98% for smear-positive TB, they have shown poorer sensitivity (66%) for smear-negative TB (Nicol and Zar, 2011, Parida and Kaufmann, 2010), and still present some notable drawbacks particularly for PoC use (LaBarre et al., 2011). These tests require trained personnel, are labour-intensive and may present risk of cross-contamination of samples (Nicol and Zar, 2011, Parida and Kaufmann, 2010). Furthermore, the performance of the NAATs is affected by the sample preparation, the extraction of DNA and the presence of PCR inhibitors (Chang et al., 2012). However, the fully automated Xpert MTB/RIF™ system offers an improved and easy-to-use NAAT alternative (Nicol and Zar, 2011, Parida and Kaufmann, 2010). The Xpert MTB/RIF™ detects MTB and rifampin resistance as a surrogatee for multi-drug resistant (MDR) TB directly from a patient's sputum in less than 2 hours with minimal risk of sample contamination (Nicol and Zar, 2011, Parida and Kaufmann, 2010). Moreover, the test has a sensitivity profile that is comparable to some culture methods but without the need for Bio-safety level 3 facilities (LaBarre et al., 2011).While the method presents prospects for near-patient care TB detection facilities, the high cost of the Xpert MTB/RIF™ system and operational challenges may be limiting factors, particularly in low income, TB-burdened populations (LaBarre et al., 2011).

Collectively, the currently available TB diagnostics present drawbacks of low sensitivity or specificity, the inability to distinguish between latent and active disease, the

requirement of sophisticated instrumentation, special storage conditions and proficient technical workforce, and high cost implications (Chamie et al., Dorman, 2010, Ernst et al., 2007, Kaufmann and Parida, 2008). Apart from the shortcomings of the current diagnostic tools, access to TB diagnostics remains a challenge (Dorman, 2010). These challenges necessitate the development of an affordable; sensitive; specific; user-friendly; rapid and robust; equipment-free and deliverable to end-users (ASSURED) TB diagnostic tools (LaBarre et al., 2011). In addition, these ASSURED TB diagnostic tools must be tailored for use in high TB and HIV prevalence developing countries (Chamie et al., Dorman, 2010, Ernst et al., 2007, Kaufmann and Parida, 2008).

1.4 Surrogate TB biomarkers and alternative tools for detection of MTB antigens

1.4.1 ESAT-6 and CFP-10 are surrogate TB biomarkers

Reliable target biomarkers are important when developing diagnostic tools (Kaufmann and Parida, 2008, Parida and Kaufmann, 2010). The early secreted antigenic target-6 (ESAT-6) and culture filtrate protein-10 (CFP-10) are considered good surrogate biomarkers for TB detection (Andersen et al., 2000, Parida and Kaufmann, 2010). ESAT-6 and CFP-10 are secreted by the ESX-1 secretion system, which is encoded for by the MTB genomic region of difference one (RD1) (Andersen et al., 2000, Parida and Kaufmann, 2010). The RD1 is crucial in the pathogenesis of MTB, and distinguishes TB from non-TB-causing mycobacteria (Andersen et al., 2000, Parida and Kaufmann, 2010). This region is deleted in all *Bacillus Calmette-Guérin* (BCG) vaccine strains of *M. bovis,* and is the primary deletion event responsible for attenuation of virulent *M. bovis* (Andersen et al., 2000, Parida and Kaufmann, 2010, Pym et al., 2002, Pym et al., 2003). The RD1 region spans open reading frames Rv3871 to Rv3879c (Andersen et al., 2000, Parida and Kaufmann, 2010). CFP-10 and ESAT-6 are encoded by Rv3874 and Rv3875, respectively, and are the two most predominant secretory proteins encoded by the RD1 region (Andersen et al., 2000, Parida and Kaufmann, 2010). However, the RD1

is also present in *M. avium*, in the facultative-pathogenic *M.kansasii*, and the non-pathogenic *M. smegmatis*. In *M. smegmatis* for instance, the region is proposed to play a role in DNA uptake (de Jonge et al., 2007), suggesting that pathogenic mycobateria have adapted an ancestral system for protein secretion that is necessary for survival and multiplication in the host cell (Andersen et al., 2000, Parida and Kaufmann, 2010).

ESAT-6 and CFP-10 are implicated in the pathogenesis of TB (Andersen et al., 2000, Ernst et al., 2007, Parida and Kaufmann, 2010). While the precise role of these proteins is not clearly defined, they elicit cellular immune response in MTB-infected individuals (Andersen et al., 2000, Ernst et al., 2007, Parida and Kaufmann, 2010). Their expression is associated with an increase in the cytolytic ability of MTB (Lewis et al., 2003). The proposed pathogenesis mechanism for ESAT-6 and CFP-10 is two-fold: T-cell activation and macrophage deactivation (de Jonge et al., 2007). A remarkable feature of the CFP-10.ESAT-6 complex is the long flexible arm formed by the C-terminus of CFP-10 (de Jonge et al., 2007). This arm is considered a contact point through which MTB interacts with host cell surfaces. Thus, due to their role in MTB pathogenicity and relatively restricted distribution in pathogenic mycobacteria, the antigenic pair may serve as good surrogate biomarkers for TB detection (Andersen et al., 2000, Parida and Kaufmann, 2010).

For instance, the use of ESAT-6 and CFP-10 in the diagnosis of TB has been demonstrated; particularly in the IGRAs, the T.SPOT.TB, and QuantiFERON –TB Gold (QTF-G), respectively (Andersen et al., 2000, de Moraes Van-Lume, 2010, Dheda et al., 2009, Ernst et al., 2007, Parida and Kaufmann, 2010). The IGRAs, T.SPOT.TB and QTF-G are serodiagnostic assays and offer rapid TB diagnosis. These assays have been developed to measure the immune response to ESAT-6 and CFP-10 (Andersen et al., 2000, de Moraes Van-Lume, 2010, Dheda et al., 2009, Ernst et al., 2007, Parida and Kaufmann, 2010). While these assays offer a single patient visit and modestly high sensitivity, they have significant limitations in effectively diagnosing active TB

(Andersen et al., 2000, de Moraes Van-Lume, 2010, Dheda et al., 2009, Ernst et al., 2007, Parida and Kaufmann, 2010). In fact, these assays are mostly effective in the detection of latent tuberculosis (Andersen et al., 2000, de Moraes Van-Lume, 2010, Dheda et al., 2009, Ernst et al., 2007, Parida and Kaufmann, 2010).

Another example of ESAT-6.CFP-10-based TB diagnosis used monoclonal antibodies against the two antigens in a sandwich ELISA (Feng et al., 2011). The performance of the ELISA was tested on sputum culture supernatants and pleural effusion aspirates (Feng et al., 2011). The ESAT-6 monoclonal antibody-based ELISA had a sensitivity of 95.4% and a specificity of 100% in TB samples; while the CFP-10 monoclonal antibody-based ELISA had a sensitivity and specificity of 81.6% and 92.2%, respectively (Feng et al., 2011). Furthermore, the results showed positive detection rates of ESAT-6 and CFP-10 of 86.8% (33/38) and 76.3% (29/38), respectively, in the clinical diagnosis of tuberculosis pleural effusion in patients that were negative for MTB culture (Feng et al., 2011). Yet, these types of antibody-based assays have drawbacks which hinder their use at PoC. For instance, in the current format, this ESAT-6.CFP-10 based detection assay requires equipment and a laboratory, and highly skilled technicians to run and analyse the test results. Besides, the patient samples have to be subjected to liquid bacterial culture prior to being tested for TB on the ESAT-6.CFP-10 based ELISA. This approach is therefore not feasible for PoC use. Nonetheless, there are some commercially available rapid tests that use antibodies as detection reagents (Pendergrast et al., 2005). However, antibody-based tests can present drawbacks such as low specificity for cognate targets and the requirement of special transportation and storage facilities (Pendergrast et al., 2005). These limitations call for alternative detection tools that will allow for immediate, sensitive and specific diagnosis of TB at PoC.

1.4.2 Aptamer-based sensors that detect ESAT.6.CFP-10 as alternative TB diagnostic tools

Aptamer-based sensors that detect pathogenic antigens and disease biomarkers such as the ESAT-6.CFP-10 heterodimer are promising alternative tools for diagnosing TB. Aptamers are an interesting class of molecules with high specificity and sensitivity to their respective targets (Famulok et al., 2007, Lee et al., 2010). These molecules, which rival antibodies in function, are aptly defined as "chemical antibodies" (Famulok et al., 2007, Lee et al., 2010). Aptamers have many of the functional properties and applications of monoclonal antibodies (Famulok et al., 2007, Lee et al., 2010). In addition, aptamers have advantages over antibodies; particularly in the diagnostics application.

1.5 Properties and prospects of aptamers

1.5.1 Properties of aptamers

Aptamers are artificial nucleic acid ligands that can bind with high affinity and specificity to their cognate targets (Lee et al., 2008). Aptamers can in principle be selected against virtually any target including non-immunogenic molecules, as the method through which they are isolated does not require the use of living organisms (Strehlitz et al., 2008). Aptamers are generated from combinatorial oligonucleotide or peptide libraries by an *in vitro* process called systematic evolution of ligands by exponential enrichment (SELEX) (Ellington and Szostak, 1990b). SELEX is an iterative process typically comprising multiple rounds of alternate cycles of ligand selection from pools of diverse sequences and amplification of target-binding species (Ellington and Szostak, 1990b). The process yields a pool of aptamer sequences that will have have formed specific structural compatibility interactions (secondary and tertiary) with the target of interest (Ellington and Szostak, 1990b).

Aptamers are typically about 22 to 100 nucleobases long, and comprise a variable region flanked by constant regions which allow for amplification and sequence identification (Ellington and Szostak, 1990b, Kanwar et al., 2011, Simmons et al., 2012). Similar to monoclonal antibodies, aptamers have a number of applications including biomarker identification (Kim et al., 2009c), therapeutics (Shigdar et al., 2011), diagnostics (Famulok et al., 2007, Khan, 2008, Savory et al., 2010, Zhu et al., 2010) and high throughput screening (Green et al., 2001). However, by virtue of their biophysical properties, aptamers rival antibodies in many aspects (Proske et al., 2005).

1.5.2 Aptamers versus antibodies

Compared to antibodies, aptamers have lower detection limits; in the low nanomolar to picomolar range (Chiu, 2009, Famulok et al., 2007, Pendergrast et al., 2005, Proske et al., 2005). Another key advantage of aptamers is their ability to fold into multiple intricate structures to specifically accommodate their respective targets (Proske et al., 2005). Consequently, aptamers can be isolated against a wide range of targets, and bind them in a perfect manner akin to that of antibody-antigen interactions (Proske et al., 2005). Aptamer targets include small molecules such as amino acids, chemical toxins, peptides, proteins, polysaccharides, and whole microorganisms such as bacteria, viruses and protozoa (Chiu, 2009, Famulok et al., 2007, Khan, 2008, Lee et al., 2010).

In contrast to antibodies, aptamers are much smaller and compact; therefore bypass challenges such as steric hindrance when interacting with their respective targets (Chiu, 2009, Famulok et al., 2007, Khan, 2008, Lee et al., 2010). In clinical applications, aptamers have faster tissue penetration and better bioavailability (Simmons et al., 2012). The non-immunogenic nature of aptamers also make them ideal therapeutic agents (Simmons et al., 2012). Furthermore, unlike antibodies, aptamers (particularly DNA-based) are more stable and able to tolerate harsh chemical, physical and biological conditions (Proske et al., 2005). Thus, aptamers maintain their functional integrity even

after exposure to extreme conditions, which would generally be destructive to antibodies (Chiu, 2009, Famulok et al., 2007, Khan, 2008, Lee et al., 2010).

In addition, aptamers have a simple conjugation chemistry, which makes them amenable to various chemical modifications for improved stability and the attachment of reporter molecules for various applications (Pendergrast et al., 2005). These properties make aptamers easily compatible with most biosensor devices (Famulok et al., 2007, Pendergrast et al., 2005, Xia et al., 2010), including multiplex systems which are able to detect different analytes in patients' samples (Chiu, 2009, Famulok et al., 2007, Pendergrast et al., 2005). Moreover, the production of aptamers is easier and less expensive relative to that of antibodies (Pendergrast et al., 2005). Following initial isolation *in vitro*, aptamers can be produced in large amounts either by enzymatic methods or chemical synthesis (Pendergrast et al., 2005). In chemical synthesis, aptamers can be produced in lyophilized form without batch to batch variation, and can be stored at room temperature for up to a year and retain full functionality activity (Pendergrast et al., 2005).

The properties and advantages discussed above; make aptamers attractive candidate probe molecules for the development of highly sensitive and selective detection methods (Kanwar et al., 2011, Pendergrast et al., 2005). Overall, aptamer-based detection systems offer such advantages as rapid detection, high sensitivity and specificity, stability of detection probe, ease of use, relatively low cost of production, and minimal technical hurdles (Famulok et al., 2007, Khan, 2008, Savory et al., 2010, Zhu et al., 2010).

1.5.3 Aptamer-target interaction

The term "aptamer" is derived from the Latin word *aptus*, meaning "to fit", and the Greek word *meros*, meaning "part" or "region" (Ellington and Szostak, 1990b, Robertson and Joyce, 1990, Tuerk and Gold, 1990). The combination of *aptus* and

meros describes the inherent manner in which an aptamer interacts with its target. These interactions occur mainly via structural compatibility between the aptamer and the target (Kanwar et al., 2011, Pendergrast et al., 2005).

Aptamer structures are driven by and formed through complementary base-pairing (Pendergrast et al., 2005). These secondary structures primarily consist of short helical (double-stranded) stems and single-stranded loops (Hermann and Patel, 2000). These stem-loop structures form the basis of aptamer-target interaction (Figure 1-1). The helical regions stabilise the secondary structures and allow the single-stranded regions (loops and such others) to "collapse" into the most likely three-dimensional shapes that are suitable for target interaction (Gold et al., 2012).

The multiple stable tertiary structures resulting from the different combinations of secondary structures, allow aptamers to bind to their respective targets via electrostatic interactions, ionic or dipole forces, Van der Waals interactions, as well as hydrogen bonds and stacking interactions (Hermann and Patel, 2000, Pendergrast et al., 2005). Sequence-specific interactions of aptamers with their respective targets are based on hydrogen bonds and Van der Waals interactions, while nonspecific interactions result from electrostatic interactions (Jones et al., 2001).

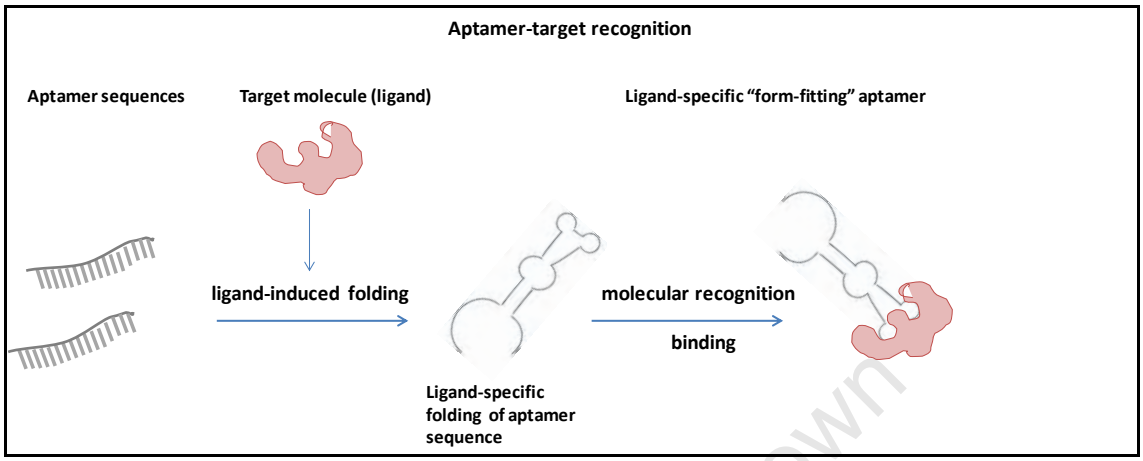


Figure 1-1: Schematic illustration of aptamer-target interaction. Upon encountering a target, an aptamer assumes a target-induced conformation. The structural compatibility is a consequence of the selection pressure from the SELEX process.

1.5.4 Applications of aptamers

Aptamers have a wide range of applications; in basic research (Marro et al., 2005), biomarker discovery, in environmental analytics (Stoltenburg et al., 2012, Kim et al., 2009a) and clinical practice (Moore et al., 2011, Ng et al., 2006, Rusconi et al., 2004). For basic research purposes, aptamers can be used for target validation and biomarker discovery (Chang et al., 2013), as well as to study the functional properties and mechanisms of complementary targets such as proteins and enzymes (Marro et al., 2005). Environmental applications of aptamers include the detection and removal of toxic substances for such purposes as water remediation. In 2009, Kim and colleagues developed DNA aptamers that are able to remove arsenic from Vietnamese groundwater (Kim et al., 2009b). Furthermore, aptamer-based sensors can be used in the identification of toxins in food and agricultural products (Kim et al., 2009b). Recently, NeoVentures Biotechnology has successfully commercialized the first aptamer based diagnostic platform for analysis of mycotoxins in grain (Rusconi et al., 2004).

In clinical practice, aptamers can be used both as detection reagents in diagnostics, and as active molecules in therapeutic agents (Rusconi et al., 2004). Macugen (Pegaptanip) is the first aptamer-based therapeutic agent to be used in clinical practice (Ng et al., 2006). Macugen is used to treat macular degeneration (Ng et al., 2006). The therapeutic is based on an RNA aptamer with both a high affinity for, and an antagonistic effect on the Vascular Endothelial Growth Factor (VEGF), and can thus inhibit angiogenesis (Ng et al., 2006). Aptamers also have potential applications in the control of infectious diseases such as HIV (Dey et al., 2005, Khati et al., 2003, Moore et al., 2011). Moreover, aptamers can be used, with suitable nanocarriers for targeted drug delivery (Kanwar et al., 2011). For instance, aptamers isolated against prostate specific membrane antigen (PSMA) was conjugated to nanoparticles for potential targeted prostate cancer therapy (Ng et al., 2006).

1.5.4.1 Aptamers in diagnostics

In clinical diagnostics, aptamers can be used as probe molecules in a variety of molecular sensors for the detection of specific biomarkers (Reviewed in (Kanwar et al., 2011). These sensors can take on a range of formats such as ELISA (Drolet et al., 1996), electrochemical devices (Ikebukuro et al., 2004), quartz crystal-based detectors (Liss et al., 2002), Surface Plasmon Resonance (SPR) technology (Proske et al., 2005, Wang et al., 2008, Xia et al., 2010), lateral flow strip assays (Liu et al., 2009, Liu et al., 2006), and simple, single step ("mix and measure") nanoparticle solution-based detection systems (Proske et al., 2005, Wang et al., 2008, Xia et al., 2010). As indicated, such properties as size, conformational flexibility, ease of modification and high stability, make aptamers readily compatible with most biosensor designs.

Both RNA and DNA aptamers can be easily conjugated to reporter molecules for use in biosensors, without compromising their functionality (Chiu, 2009, Famulok et al., 2007, Khan, 2008, Lee et al., 2010). However, when developing aptamer-based sensors, DNA

aptamers are usually preferable due to the inherent stability of DNA relative to RNA (Moore et al., 2011, Neves et al., 2010). Generally, DNA aptamers can be readily synthesized, chemically modified, refolded, and integrated into microarrays, microfluidics, sandwich assays and electrochemical biosensors (Neves et al., 2010). Examples of DNA-aptamer based biosensors include those that detect ATP (Jhaveri et al., 2000), thrombin (Li et al., 2002), cocaine (Stojanovic et al., 2001), PDGF (Fang et al., 2003), Type 2 Diabetes biomarker (Retinol Binding Protein 4) (Lee et al., 2008) and avian influenza virus, H5N1 (Bai et al., 2012).

1.5.4.1.1 Simple aptamer-based detection systems: potential design suitable for PoC diagnostic

In the construction of aptamer-based biosensors, several methods have been developed to translate the aptamer-target binding events into physically detectable signals (Wang et al., 2008). These methods include functionalizing aptamers for optical or electrochemical sensing with fluorophores, nanoparticle labels or label-free detection systems (Wang et al., 2008). Gold nanoparticles (AuNPs) have been widely used to enable simple visual detection in diagnostics kits and immunochromatographic strips (Wang et al., 2008, Xia et al., 2010, Xu et al., 2009). This is due to the unique optical properties of AuNPs, which allow for sensitive detection of molecules from an analyte within minutes (Wang et al., 2008, Xia et al., 2010, Xu et al., 2009). Other beneficial properties of AuNPs include their long-term stability and compatibility with most biomolecules including antigens, antibodies, and DNA (Wang et al., 2008, Xia et al., 2010, Xu et al., 2009).

The use of aptamer-AuNPs conjugates for improved, highly sensitive and specific molecular diagnostics with low technical hurdles has been demonstrated (Chiu, 2009, Xia et al., 2010). Examples of aptamer-AuNP detection systems include lateral flow (Liu et al., 2009, Liu et al., 2006) and colloidal gold solution sensors (Chiu, 2009, Liu and Lu, 2006, Xia et al., 2010). These aptamer-AuNP solution based detection systems

have been shown to work for the detection of both recombinant targets in buffer and native forms of the target in biological fluids such as serum (Chiu, 2009, Xia et al., 2010). The principle of an Aptamer-AuNP "one-step" detection system is illustrated (Figure 1-2).

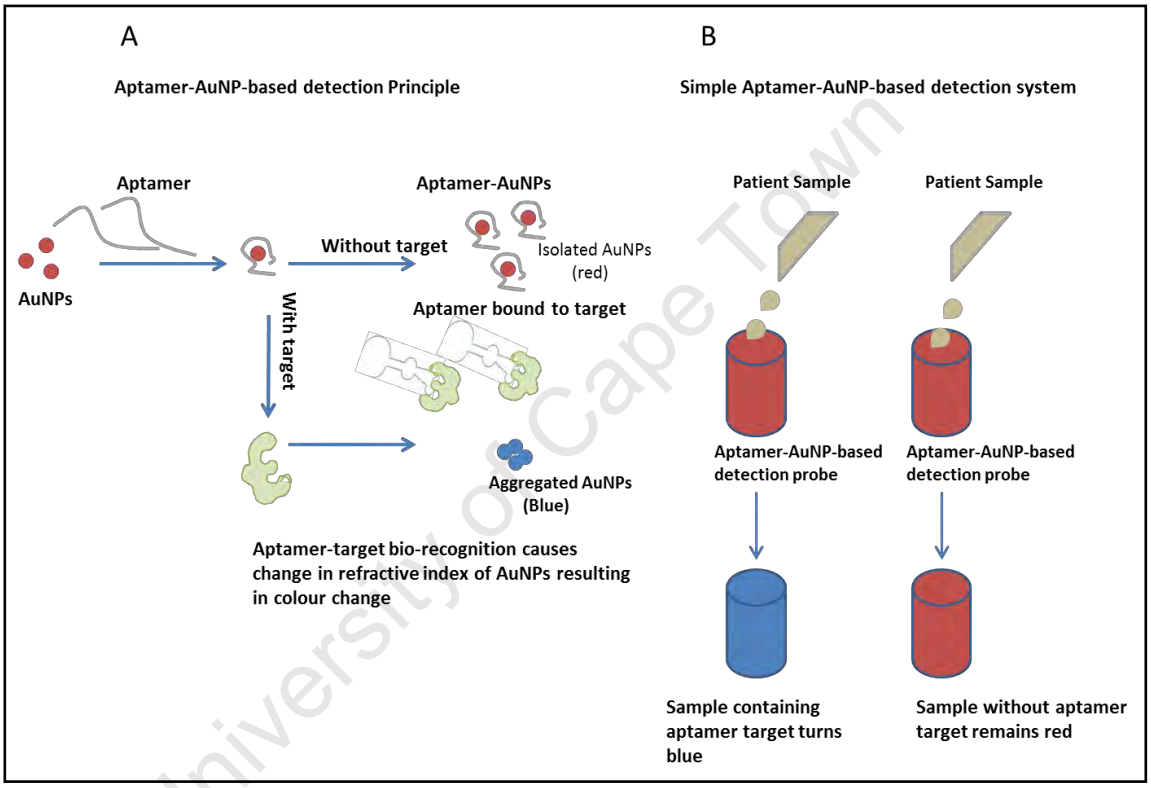


Figure 1-2: Schematic illustration of a simple aptamer-AuNPs-based detection system. A. The principle of aptamer-AuNP-based target recognition. Aptamers are conjugated to AuNPs through various methods to form simple detection probes. The conjugated aptamer prevents aggregation of the AuNPs. The distance between the AuNPs affects their optical properties and colour. Upon target-binding, the aptamers are sequestered from the AuNPs, shortening the distance between particles and causing a change in the refractive index which results in colour change. B. Schematic illustration of a potential simple "mix and detect" Aptamer-AuNPs based detection system. Addition of the patient sample containing the aptamer target will change colour in the detection probe, while the sample that does not contain the aptamer target will not change colour.

1.5.5 Aptamer characterisation: approaches for determining aptamer fitness for use

There are certain criteria that aptamer-based detection probes have to meet in order to be incorporated into a sensing platform for biosensor design (Famulok et al., 2007). These criteria include high affinity and specificity for the target, as well as appropriate sequence length for optimal functionality (Famulok et al., 2007, Proske et al., 2005). Several methods have been employed in characterising aptamers on the basis of these criteria. Common methods include quantifying aptamer-target interactions by determining their binding kinetics (Murphy et al., 2003, Ruigrok et al., 2012, Win et al., 2006), and analysing the structural properties of the aptamers to determine the key features that contribute to their target-recognition function (Nonaka et al., 2010, Shangguan and Tang Z., 2007, Wuchty S, 1998, Zhou, 2010).

1.5.5.1 Functional characterisation and binding affinities of aptamers

An assay similar to ELISA is commonly used to determine binding of aptamers to their respective targets (Drolet et al., 1996). In this assay, termed enzyme linked oligonucleotide assay (ELONA), aptamers replace antibodies as target recognition molecules (Drolet et al., 1996). The ELONA generally works for an end-point read-out, providing a "yes/no" answer, suitable for high throughput screening of individual aptamers clones for target-recognition ability. In order to study aptamer-target interactions in more detail, and to determine the binding affinities or kinetics, technologies such as Electrophoretic Mobility Shift Assay (EMSA) and Surface Plasmon Resonance (SPR) can be used for semi-quantitative and quantitative analysis, respectively (Murphy et al., 2003, Win et al., 2006).

EMSA can be used for the characterisation of biomolecular interactions in various qualitative and quantitative analyses, particularly for detecting protein-nucleic acid complexes (Hellman and Fried, 2007). Typically, solutions of proteins and radiolabelled

nucleic acids are combined, and subjected to native or non-denaturing electrophoresis using either polyacrylamide or agarose gel (Hellman and Fried, 2007). Following electrophoresis, the distribution and migration patterns of the species containing nucleic acid throughout the gel is determined, usually by autoradiography or phosphorimaging (Hellman and Fried, 2007). The general principle is that during electrophoresis, the protein-nucleic acid complexes migrate more slowly than the corresponding free nucleic acids; therefore aptamer-target complexes versus free (unbound) aptamer can be visualized and quantified. While EMSA is traditionally used to assess binding and conformational characteristics of protein-DNA interactions, it can under specific conditions be used to determine binding stoichiometry (based on the amount of complex formed as a function of protein concentration at equilibrium), as well as kinetic parameters such as ratio of the association and dissociation constants or (Hellman and Fried, 2007, Demarse et al., 2009).

Conversely, molecular interactions between aptamers and cognate ligands can be studied in real-time and quantified using SPR technology (Luzi et al., 2003). SPR is a reference optical method which allows for the label-free detection and quantification of specific ligand-analyte interactions in real time (Luzi et al., 2003). The technology enables rapid, high-throughput screening of molecular interactions and associated equilibrium and kinetic binding properties (Li et al., 2006, Potyrailo et al., 1998, Tombelli et al., 2005, Win et al., 2006). As a result, the technology has been widely employed to determine interactions between a range of aptamers and their respective targets (Li et al., 2006, Potyrailo et al., 1998, Tombelli et al., 2005, Win et al., 2006). Essentially, SPR can be used to determine whether or not aptamer-target molecules interact, how fast or how slow they dissociate, and in some cases, if there are any conformational changes associated with these interactions (Luzi et al., 2003). This technology shows some advantages over techniques such as EMSA in that the

interaction is probed in real-time, and the information required for quantifying kinetic parameters is easily obtained (Luzi et al., 2003).

In a typical SPR experiment, a ligand is immobilised on a metal coated sensor chip surface, followed by injection of a sample solution containing the analyte at a constant flow rate (Win et al., 2006). The detector part of the SPR instrument then monitors and records the binding events as the change in the refractive index (resonance angle occurring at the chip surface) as a function of time (Luzi et al., 2003). This change in refractive index angle is proportional to the amount of bound molecules (Luzi et al., 2003). The sensing signal is then recorded relative to the base signal, and gives an indication of the strength of the interaction (Luzi et al., 2003). From this output, key kinetic parameters such as the association (ka), dissociation (kd), and equilibrium (K_D) constants can be derived (Luzi et al., 2003). This information then gives an idea of the likelihood of the aptamer-target complex to associate or dissociate reversibly into its composite parts (Gopinath, 2010). Overall, SPR allows for the convenient quantitative characterisation of aptamer-target interactions (Win et al., 2006). In addition, the high-throughput nature of the SPR platform may be extended to screening the optimal truncated and modified versions of aptamers, thereby circumventing the need for the more time-consuming and labour-intensive chemical probing experiments and the occupational hazard associated with radioactive shift assays (Win et al., 2006).

1.5.5.2 Computational characterisation of aptamer secondary structures

In contrast to protein folding, secondary structures of nucleic acid molecules such as aptamers provide information that fairly represents their thermodynamics and folding kinetics (Wolfinger et al., 2004, Wuchty S, 1998, Michael T. Wolfinger et al., 2004). Consequently, secondary structures have been used as a "surrogate for function" (Cowperthwaite and Ellington, 2008) for nucleic acids. Likewise, secondary structures of aptamers can provide a level of description that adequately depicts aptamer-ligand interactions (Patzel, 2004). As previously mentioned, aptamer-target recognition is

driven by structural compatibility interactions. These interactions are primarily a consequence of aptamer secondary structures (Patzel, 2004). This concept is also demonstrated by *in silico* selections, which are secondary-structure-based (Hall et al., 2007, Patzel, 2004), as well the post-SELEX analyses and optimisation of aptamer structures (Cowperthwaite and Ellington, 2008, Nonaka et al., 2010).

Secondary structures of nucleic acid aptamers can be represented in a number of different conformations consisting of helical stems (double stranded regions—with two or more paired bases) and single stranded loops (Flamm et al., 2004, Wuchty S, 1998). The helices are typically composed of Watson-Crick (A-U/T, T/U-A, C-G, G-C) or wobble (G-U, U-G) base pairs (Flamm et al., 2004, Wuchty S, 1998). The overall structure can be decomposed into a collection of different configurations of these stacked helical stems and non-paired bases (Figure 1-3), the arrangement of which is specifically defined by the cognate target.

Secondary structures of aptamers are stabilised in part by hydrogen bonds between the canonical bases (Flamm et al., 2004, Wuchty S, 1998), and stacking interactions between the aromatic rings (Mignon et al., 2005). The stacking interactions form between base pairs in helical stem structures, irrespective of the actual nucleobase (Flamm et al., 2004, Wuchty S, 1998). Furthermore, the overall structure of the aptamer is stabilised by tertiary interactions, however to a lesser extent (Ellington and Szostak, 1990a). These tertiary interactions can include pseudo-knots, non-standard base pairings and base-triplets (Ellington and Szostak, 1990a, Leontis et al., 2006). Another common conformational arrangement that can be assumed by aptamers is the G-quadruplex. This is a planar conformation constituted by an array of four guanines, with each guanine paring two neighbours via Hoogsteen bonding (Nonaka et al., 2010, Potty, 2008). This conformation is known to present highly stable and robust structure and well defined contact points for aptamers (Nonaka et al., 2010, Potty, 2008). All these interactions influence the folding of aptamers into multiple target-specific conformations (Jayasena,

1999). Like any nucleic acid, the number of potential structures that an aptamer sequence is capable of forming (N) is hypothesised to increase in a roughly exponential manner as a function of length (L) (Ellington and Szostak, 1990a, Jayasena, 1999), as depicted by the equation: $N = (1.8)^{L}$ (Ellington and Szostak, 1990a, Jayasena, 1999, Zuker and Sankoff, 1984).

Secondary structure folding is an energy-driven process whose goal is to achieve maximum complementary nucleotide base-pair stacking at the lowest possible energy cost (Heine et al., 2006, Wuchty S, 1998). The extent to which an aptamer secondary structure is able to fulfil these requirements, gives an indication of the stability of that particular structure (Flamm et al., 2000, Flamm et al., 2004). During the folding process, each assembly of an energetically favourable helical (double-stranded) region is complemented with the formation of an energetically unfavourable loop (unpaired) region (Flamm et al., 2004, Heine et al., 2006, Wuchty S, 1998). This tension of "frustrated" energetic patterns (Wuchty S, 1998) is responsible for the multiple variations of helix-and-loop arrangements for any individual nucleic acid sequence (Flamm et al., 2000, Heine et al., 2006).

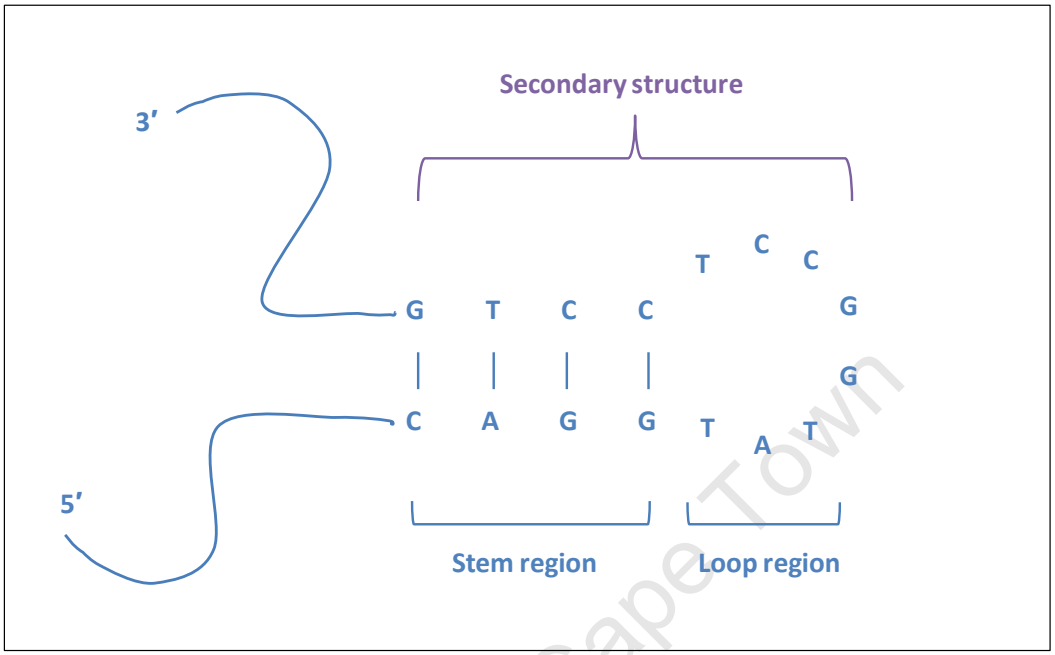


Figure 1-3: Schematic illustration of typical secondary structure composition of nucleic acid aptamers. The structures generally consist of unfolded regions (solid lines), double stranded (helical) stem regions (formed by complementary nucleobase-pairing), and single stranded loop regions. The stem regions help stabilize the overall structure while the loops (also known as the "faces") form specific contact points with cognate ligands.

Secondary structures provide not only the geometric skeleton, but also the thermodynamic and kinetic framework for the full tertiary structures of nucleic acids (Heine et al., 2006, Wuchty S, 1998). The folding energies of secondary structures are known to account for a significant portion of the overall folding energy of the full tertiary structures of nucleic acids (Flamm et al., 2000). Hence, examining secondary structures of aptamers can provide such analogous information; which will correspond with or describe (at least in part) the structural features of aptamers that influence their target-recognition ability (Heine et al., 2006, Wuchty S, 1998).

The secondary structure folding process has to adhere to a specific set of rules. The first is to maximise the number of admissible or "legal" base pairs (Flamm et al., 2000).

Here, the basic structural building block is an individual base pair that follows either the Watson-Crick or non-Watson-Crick rules outlined above (Wuchty S, 1998). The second obligation is to reduce the energy cost of the secondary structure by minimising its free energy (Flamm et al., 2000, Flamm et al., 2004). The lower the overall free energy, the more stable the structure (Flamm et al., 2000, Flamm et al., 2004). Secondary structures of nucleic acids, and by extension, of aptamers, can be adequately predicted using computational modelling methods (Aviram et al., 2012, Lorenz et al., 2011, Mathews and Turner, 2006, Zuker, 2003).

Nucleic acid secondary structure and hybridisation prediction tools have been developed using dynamic programming algorithms (Flamm et al., 2000, Lorenz et al., 2011, Zuker, 2003). These algorithms have been refined over the years based on desired applications; particularly the need to answer specific structural questions and represent respective biological systems more adequately. The most common of these secondary structure prediction tools are mfold (Zuker, 2003), and the Vienna RNA package (reviewed in (Gruber et al., 2008)). These programs use standard energy models of biopolymers (e.g. nucleic acids) in solution (Flamm et al., 2000), which account for the thermodynamic parameters for several motifs typically composite of secondary structures (Wolfinger et al., 2004). These motifs include Watson-Crick base pairs, internal mismatches, terminal mismatches, terminal dangling ends, hairpins, bulges, internal loops, and multibranched loops (reviewed in (SantaLucia and Hicks, 2004)). The common standard energy model used by these prediction programs defines the overall free energy of a given nucleic acid secondary structure as the sum of sequence-dependent contributions for each loop (Flamm et al., 2000). The secondary structure modelling thereof typically results in a single structure which reasonably meets the above mentioned criteria for nucleic acid folding (i.e. legal base pairing and low free energy). The resultant structure, usually the one most likely to be stably assumed, is termed the minimum free energy (MFE) conformation (Flamm et al., 2000).

INTRODUCTION

It is well known that the folding of nucleic acids into secondary structures is not a static process (Flamm et al., 2000, Zuker and Sankoff, 1984). Moreover, this dynamic process does not always result in a single structure for any given sequence (Flamm et al., 2000, Zuker and Sankoff, 1984). While the MFE conformation is often the most common and dominant structure occurring at equilibrium (Flamm et al., 2000), it is usually not the only structure that can occur (Wolfinger et al., 2004, Wuchty et al., 1998). In fact, with the formation of the MFE for any given nucleic acid sequence, comes a subset of suboptimal structures with low free energies within the same range as that of the MFE (Wuchty et al., 1998, Zuker and Sankoff, 1984). These suboptimal structures represent well-defined transition steps encountered along the folding pathway that eventually results in the MFE (Wolfinger et al., 2004). These transitional steps are essentially alternative conformations to the MFE, which the nucleic acid sequence can also stably assume (Wolfinger et al., 2004). Hence, these alternative conformations may in some instances act as "kinetic traps" (Wolfinger et al., 2004) which hinder equilibration, i.e. prevent the folding of a sequence into a "true" and stable MFE within the typical lifetime of that nucleic acid molecule (Flamm et al., 2000, Heine et al., 2006, Wuchty et al., 1998). Furthermore, these intermediate structures can affect the biological function of the nucleic acid, as it is possible for alternative conformations of the same nucleic acid sequence to perform completely different functions (Wolfinger et al., 2004).

Congruently, secondary structures of aptamers are defined by similar folding dynamics as those described above. Along the folding path from the open chain to the MFE, aptamers can also fold into a range of meta-stable structures, which could compete with the MFE and interfere with its functionality (Gruber et al., 2008, SantaLucia and Hicks, 2004, Wolfinger et al., 2004). Essentially, aptamers can fold into target-binding structures that are deemed "ligand-competent", as well as those that are "non-ligand-competent" (Hall et al., 2007, Mathews and Turner, 2006). In the event that the former docks the cognate target, a stable target-specific conformation will form (Ellington and

Szostak, 1990a, Jayasena, 1999). This process is time-dependent; like any reaction proceeding towards a state of equilibrium (Hall et al., 2007, Mathews and Turner, 2006). This means that there is a specific time frame within which reactants (in this case aptamer and cognate ligand) are consumed and products (aptamer-ligand complexes) are formed (Hall et al., 2007, Mathews and Turner, 2006, Patzel, 2004). If the ligand is not bound to the aptamer within this time frame, proper aptamer folding may not occur, thus resulting in non-optimal aptamer function (Hall et al., 2007, Heine et al., 2006, Mathews and Turner, 2006, Patzel, 2004). Hence, a detailed analysis of the folding dynamics of aptamer secondary structures may help characterise the target-recognition fitness of certain aptamer structures. This type of analysis will give a more dynamic view of the aptamer folding process, beyond the "static" picture provided by standard energy minimisation methods, which only present the end-point conformation (i.e. the optimal (MFE) and near-optimal secondary structures). To better describe the dynamics of such a folding process, folding landscape modelling is used (Cowperthwaite and Ellington, 2008).

1.5.5.2.1 Secondary structure folding landscapes: concept and significance

A folding landscape can be defined as a map of the folding process which provides information about the thermodynamics and folding kinetics of nucleic acid structures (Cowperthwaite and Ellington, 2008, Wolfinger et al., 2004). Thermodynamically, a folding landscape is defined by the ensemble of free energies of all possible conformations that a particular sequence can assume (Wolfinger et al., 2004). Each conformation is associated with both an energy value and a distance measure relative to all other possible conformations for that sequence (Cowperthwaite and Ellington, 2008, Wolfinger et al., 2004). On the other hand, folding kinetics are concerned with the change in molecule structure over time (Cowperthwaite and Ellington, 2008), which is an energy-driven process (Flamm et al., 2000, Wuchty et al., 1998). These kinetics describe the frequency with which certain nucleotides meet in the process of forming

secondary structures, and the rate of occurrence of these structures within a given time frame (Cowperthwaite and Ellington, 2008).

Folding landscapes are interesting to examine as they give an indication of nucleic acid structural fitness and stability (Cowperthwaite and Ellington, 2008, Flamm et al., 2000). They provide a compact picture of the likelihood of a conversion between the MFE and suboptimal structures of each sequence (Cowperthwaite and Ellington, 2008). The "kinetic trapping" effect triggered by the suboptimal structures encountered along the pathway to the MFE is a consequence of the energy barriers associated with these structures (Flamm et al., 2000, Heine et al., 2006, Wolfinger et al., 2004). These energy barriers represent the amount of activation energy that is required for a sequence to change from one conformation to another (Heine et al., 2006).

On the whole, folding landscapes can help determine a few key features about the secondary structures of aptamers. Firstly, they give an estimate measure of how long it takes for the system (the folding process) to reach an equilibrium state, having started from an open chain conformation (Flamm et al., 2000, Heine et al., 2006, Wolfinger et al., 2004). Secondly, they provide a visual depiction of the extent to which the intermediate (suboptimal) structures are being populated on the path from the initial structure to the MFE (Flamm et al., 2000, Heine et al., 2006, Wolfinger et al., 2004). Finally, folding landscapes can estimate the passage times (folding or refolding times) (Wolfinger et al., 2004) of specific conformations along the folding pathway. Therefore, analysis of folding landscapes of aptamers can help predict the structure that is most likely to occur without getting trapped in suboptimal intermediate conformations (Aviram et al., 2012, Flamm et al., 2000, Wolfinger et al., 2004). Furthermore, this analysis can determine how smoothly the folding occurs in time, and discriminate those structures that get stuck in meta-stable conformations within a given folding time (Aviram et al., 2012, Flamm et al., 2000); effectively characterising the structural fitness of aptamers.

1.5.5.2.2 Computational estimation of folding energy landscapes

Folding landscapes have been commonly computed using the Boltzmann probability of the MFE structure (Ancel and Fontana, 2000, Mathews and Turner, 2006). The Boltzmann probability can be defined as an approximate measure of the fraction of time that an aptamer sequence spends folded into a particular shape (Ancel and Fontana, 2000). Generally, the MFE structure will likely have the largest Boltzmann probability of any shape an aptamer sequence may assume (Mathews and Turner, 2006). If the Boltzmann probability of an aptamer's MFE is large, then the aptamer is expected to stably assume its MFE conformation most of the time (Ancel and Fontana, 2000). However, this probability can be small, in which case the aptamer is not expected to stably fold into its MFE shape (Ancel and Fontana, 2000).

Computational estimation of secondary structure folding energetics and mechanics can indeed be made possible by efficient polynomial algorithms (Wuchty et al., 1998), and allow for a more detailed analysis of the folding landscape (Wolfinger et al., 2004). These algorithms enable the efficient prediction of the MFE structure, all possible base-pairing probabilities, and all the suboptimal structures occurring within a given energy threshold from the MFE (Flamm et al., 2000, Slutsky and Mirny, 2004). Furthermore, these refined algorithms are able calculate the probability of the formation of a given structure within an ensemble of potential structures (Flamm et al., 2000, Slutsky and Mirny, 2004). However, it is important to note that molecular energies cannot be computed precisely due to the imperfect nature of the theoretical models used by these algorithms (Wuchty et al., 1998), which do not account for solvation or intermolecular influences (Flamm et al., 2000). Although, factoring in the folding energies of a sample of intermediate structures (minima) within a given range above the MFE is considered sufficient for most folding landscape estimations (Slutsky and Mirny, 2004).

A smooth folding landscape is one that has fewer suboptimal structures within a given energy ensemble and in which MFE occurs near the native state (Wolfinger et al., 2004).

This phenomenon is called "downhill folding" (Wolfinger et al., 2004), where there is little or no energy barrier, and folding into the MFE occurs relatively quickly (Heine et al., 2006). Conversely, "rougher landscapes" have at least one or more energy barriers (Wolfinger et al., 2004), which slow down transition times and adversely affect pathways to the MFE structure (Heine et al., 2006).

Alternative to the complex mathematical calculations and simulations, barrier trees can be used to describe the folding dynamics of an aptamer (Wolfinger et al., 2004). Barrier trees provide a coarse-grained simplified model of an energy landscape by discarding all (fully) unstable structures and retaining only the meta-stable conformations along the folding pathway (Schuster, 2009). In the simplest case, transition rates between these structures are modelled and presented according to their respective energy barrier heights, which represent the amount of activation energy required to switch from one conformation to another (Flamm et al., 2000). Barrier trees can be computed using the Vienna RNA *Barriers* Program (Flamm et al., 2000). The program makes it possible to sample suboptimal structures from an ensemble of possible structural motifs, and determine their equilibrium probability (Hofacker et al., 2010). The *Barriers* algorithm works by processing the potential conformations in an energy-sorted order (Flamm et al., 2000). From this "energy or cost-sorted" list of conformations, a barrier tree is constructed (Flamm et al., 2000),which contains either all possible conformations for a given sequence, or only those below a set energy threshold (Hofacker et al., 2010). While the model is simplistic in that it only represents transition rates between intermediate structures, which are represented by their respective barrier heights in the tree conformations (Slutsky and Mirny, 2004, , Heine et al., 2006, Wolfinger et al., 2004), it provides sufficient information which sums up the key features of a folding landscape (Wolfinger et al., 2004). Taken together, the secondary structure features discussed above (thermodynamics and kinetics) are worth examining in aptamer characterisation, as they collectively provide a scaffold for the potential structures of

aptamers and their influence on aptamer-target interactions. Furthermore, the bioinformatics approaches described above can be used to complement and guide laboratory-based experiments in the process of aptamer optimisation, and the subsequent design of the aptamer-based detection probes.

1.5.6 Post-SELEX Optimisation of Aptamers

The main prerequisites for aptamers as diagnostic probes are high affinity, selectivity to targets (Potty, 2008, Shangguan and Tang Z., 2007), as well as satisfactory stability in biological fluids (Dey et al., 2005, Savory et al., 2010, Shangguan and Tang Z., 2007). A key factor in aptamer-target recognition and interaction is aptamer stability, which depends largely on the biophysical properties of aptamers such as the sequence length and folding conformation (Nonaka et al., 2010, Potty, 2008). Optimisation of these features may result in improved aptamer functionality.

1.5.6.1 Aptamer sequence length, contribution to function, and optimisation by truncation

Full length aptamers, that is, those comprising a full random region (as evolved from the original SELEX library), flanked by constant regions, generally consist of two distinct motifs; target-binding/binding-facilitating motifs, and regions that do not participate in either aspect of target binding (Zhou, 2010). The elimination of the latter (i.e. non-essential nucleotides) through the post-SELEX truncation of aptamer sequences has been shown in some instances, to improve the binding functionality and kinetics of aptamers (Nonaka et al., 2010, Savory et al., 2010, Shangguan and Tang Z., 2007, Zhou, 2010).

As discussed previously, an aptamer structure takes on a target-induced conformation with numerous discriminatory intermolecular contacts that are target-specific (Flamm et al., 2000, Heine et al., 2006, Wuchty et al., 1998). On this basis, Nonaka et al (2010) hypothesised that the affinity of the aptamer should be related to the accessibility of the

target molecule to its specific binding pocket on the aptamer (Nonaka et al., 2010). An additional postulation in this regard was that reducing the length and consequently the overall structure of the aptamer, should make the target-binding domain of the aptamer more accessible, and thus improve binding affinity. Accordingly, Nonaka and colleagues set out to minimise the sequence of an aptamer isolated against the vascular endothelial growth factor (VEGF) with a view to enhance its affinity for VEGF (Nonaka et al., 2010). This modification resulted in retained affinity for the VEGF protein, and improved kinetics; from 20 nM to 1.4 nM (Nonaka et al., 2010).

From a structural point of view, truncating the aptamer sequences may improve the stability of the aptamers by favouring their ligand-competent (target-biding) structures (Ancel and Fontana, 2000, Mathews and Turner, 2006). Shortening the sequences will likely reduce the number of potential structures, given that this number is said to increase exponentially as a function of length (Zuker and Sankoff, 1984). This may decrease the likelihood that the aptamer will be trapped in suboptimal conformations in the folding process towards the optimal target-binding shape (Ancel and Fontana, 2000, Wuchty et al., 1998, Zuker and Sankoff, 1984). In addition, smaller-sized aptamers allow for more desirable dimensions when constructing simple, compact, and highly sensitive and specific biosensors (Zhou, 2010). Moreover, the production of shorter-sequence aptamers is easier and more cost effective (Shangguan and Tang Z., 2007). Accordingly, it is beneficial to determine and functionalise the minimal binding motifs of aptamers, particularly for diagnostic purposes.

1.5.6.2 Approaches to aptamer optimisation by truncation

The commonly used methods of determining the target-binding motifs of aptamers include enzymatic footprinting, partial hydrolysis, and in the case of RNA, *in vitro* transcription of the truncated sequence (Bing et al., 2010, Nonaka et al., 2010). With partial hydrolysis, multiple “mapping” attempts would have to be made and the predicted minimal sequences would then have to be produced, either enzymatically or

by chemical synthesis, before their binding capacity can be confirmed (Nonaka et al., 2010). These approaches are often either too complex or time-consuming (Nonaka et al., 2010). For instance, radioactive labelling is needed during enzymatic footprinting in order to detect the aptamer fragments, and the cost for chemical synthesis of multiple versions of shorter sequences (generated from partial hydrolysis) for "trial and error" binding analyses would be expensive (Bing et al., 2010, Rockey et al., 2011). To circumvent some of these challenges, several groups have made use of computationally predicted secondary structures to deduce the target-binding motifs of aptamers and eliminate non-essential nucleotides (Nonaka et al., 2010, Rockey et al., 2011). This approach (Figure 1-4) termed “rational truncation (Rockey et al., 2011), eliminates the complexity and cost limitations of the “wet lab” experimental investigations mentioned above. The resultant shorter and potentially structurally fitter aptamers can then be appropriately modified and incorporated into sensing platforms, for the design of simple, equipment-free, highly sensitive and specific, rapid and robust detection devices.

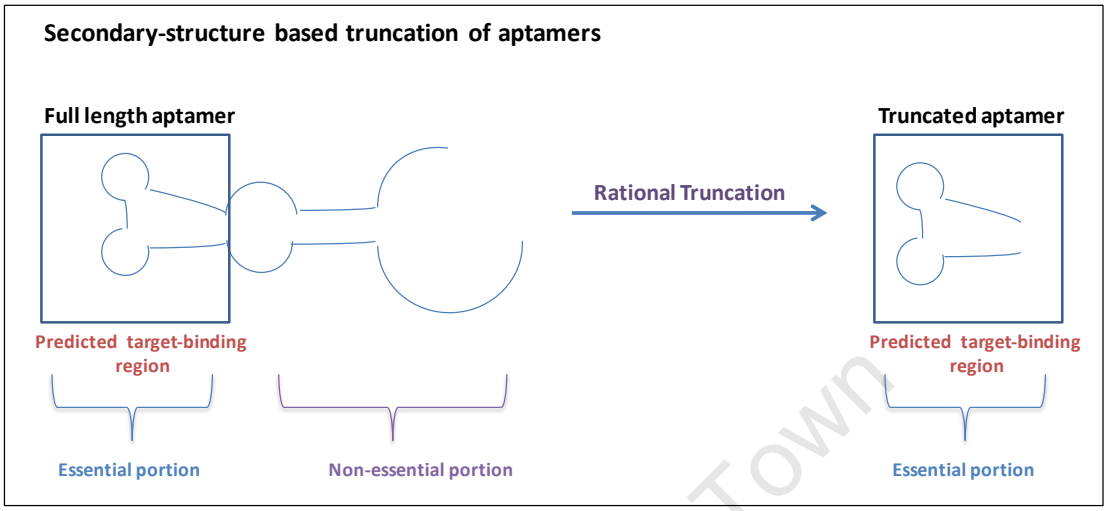


Figure 1-4: Schematic illustration of optimisation of aptamers by rational (secondary structure-guided) truncation. Rational truncation of the original full length sequences is carried out based on the predicted secondary structures, by removing non-essential parts of the sequence and retaining the predicted ligand-binding motifs. The truncated aptamer sequence is the folded into potential secondary structure to confirm the retention of the ligand-binding structure.

1.5.7 Aptamers isolated against the ESAT-6.CFP-10 heterodimer as potential TB detection probes

With the aim of exploiting the numerous advantages of aptamers and to generate detection probes for the design of a potential PoC TB diagnostic tools, recently published work from our research group at the CSIR, yielded ssDNA aptamers against the ESAT-6.CFP-10 heterodimer (Rotherham et al., 2012). The aptamers were isolated and screened for binding to the recombinant target proteins using the ELONA method previously described (Drolet et al., 1996). These aptamers showed affinity for the ESAT-6.CFP-10 heterodimer, as well as the CFP-10, and did not bind to the ESAT-6 monomer (Rotherham et al., 2012). Furthermore, one of these aptamers called CSIR 2.11 detected active TB in clinical sputum samples in ELONA readout format, with

100% sensitivity and 70% specificity, demonstrating potential clinical applications of these aptamers (Rotherham et al., 2012).

Therefore, to help design a potentially effective aptamer-based PoC TB diagnostic kit, characterisation and post-SELEX optimisation of the anti-ESAT-6.CFP-10 aptamers was necessary. Characterisation would help determine the affinity and fitness of these aptamers as detection reagents. Furthermore, describing the secondary structural properties and associated energetics of the aptamers would give insight into what structural elements are primarily involved in their functionality, and how efficiently these aptamers fold into ligand-competent structures. Collectively, this information can guide the process of developing the anti-ESAT-6.CFP-10 aptamers as probes suitable for use in a simple, rapid and highly sensitive PoC TB diagnostic kit.

1.6 Study aim and objectives

The aim of this study was three-fold: (a) to characterise the interactions of the DNA aptamers previously isolated by our research group against the ESAT-6.CFP-10 heterodimer (Rotherham et al., 2012); (b) to conduct the post-SELEX optimisation of these aptamers by rational truncation; and (c) to evaluate the functional performance of the truncated aptamers in comparison to their parental molecules.

The specific objectives were:

- To characterise the binding of six anti-ESAT-6.CFP-10 aptamers by investigating their binding kinetics and structural properties.
- To design truncated derivatives (based on predicted structural properties) of the aptamers, that would either bind similarly to, or better than the original parent sequences.
- To assess the binding functionality of the truncated aptamers as potential TB detection probes.

CHAPTER 2

METHODS

2.1 Ethics Statement

No ethics approval was required for this study as neither human samples nor animal models were used.

2.2 Bibliography Formatting and Referencing

The Harvard referencing style was used in the current study.

2.3 Protein Technology

Prior to characterisation of aptamer-protein interactions, recombinant CFP-10 and ESAT-6 proteins were expressed and purified. Each protein was expressed separately as a monomer before *in vitro* dimerization was carried out to produce the ESAT-6.CFP-10 heterodimer as previously described (Renshaw et al., 2002).

2.3.1 Expression and purification of CFP-10

The protein expression and purification protocols used in this study were performed as previously described (Renshaw et al., 2002). Briefly, primary cultures were set up using glycerol stocks previously prepared by Lia Rotherham (Rotherham et al., 2012). The glycerol stocks consisted of *E .coli* BL21 (DE3) cells (Novagen, Merck) transformed with the T7 promoter-based expression plasmid, pMRLB46 (Megan Lucas, Colorado State University) expressing full-length CFP-10 fused to an N-terminal hexahistidine tag.

2.3.1.1 Expression of recombinant CFP-10

To prepare the starter cultures for CFP-10 expression, 25mL of Luria-Bertani (LB) broth culture medium (Sigma) supplemented with 100 µg/ml ampicillin (Sigma) were inoculated with a glycerol stock. The bacterial cultures were grown overnight at 37°C in a shaking incubator at 200 revolutions per minute (rpm). The overnight starter cultures were then inoculated into 2-L Erlenmeyer flasks, containing 1 L of fresh LB medium and 100 µg/ml ampicillin, and left to grow at 37°C in a shaking incubator at 200 rpm. CFP-10 expression was induced at mid-log (exponential) phase (OD 0.6, at 600 nm), by the addition of isopropyl-1-thio-β-D-galactopyranoside (IPTG; Sigma) at a final concentration of 0.4 mM. The CFP-10 cultures were harvested 4 hours after induction by centrifugation at 6500 rpm (~7800×g) for 20 minutes at 4°C. The cell pellets were stored at -80°C prior to purification.

2.3.1.2 Purification of recombinant CFP-10

The cell pellets were thawed and subsequently lysed by the addition of 10 ml of a 1:10 Bacterial Protein Extraction Reagent (B-PER) (Thermo Scientific) diluted in ultrapure water. The B-PER is a lysis buffer consisting of a non-ionic detergent in 20mM Tris-HCl, pH 7.5; lysozyme; and DNase I. The use of the B-PER reagent enabled mild extraction of proteins from bacterial cells without the need for mechanical disruption methods such as sonication. The B-PER reagent can be used for both soluble protein extraction and inclusion body purification from bacterial cell lysates. The lysed cell suspension was homogenised using a pipette, and incubated for 15 minutes at room temperature. Thereafter, the lysate was centrifuged at 15,000 × g for 15 minutes at 4°C to separate soluble from insoluble proteins. To capture the His-tagged CFP-10, the supernatant (cytosol fraction) was loaded onto a nickel-nitrilotriacetic acid (Ni-NTA) resin column (Bio-Rad), which had been pre-equilibrated with Buffer A (6 M urea [Sigma], 25 mM Tris-HCl [Sigma], 200 mM NaCl [Sigma], 10 mM imidazole, pH 7.4, [Merck]) by moderate rotation overnight at 4°C. The resin was washed six times with 10 ml of Buffer A, with a gradual decrease in urea concentration; from 6 M to 1 M,

respectively. The column was then subjected to two washes with 20 mL of Buffer B (25 mM Tris-HCl [Sigma], 200 mM NaCl [Sigma], 10 mM imidazole [Merck], pH 7.4). The washes were collected at each step, and stored at 4°C for gel analysis. Finally, the His-tagged CFP-10 was eluted in a single step with 10 ml of Buffer B containing 300 mM imidazole. The eluate was collected, divided into 1 ml aliquots and stored at -80°C.

2.3.2 Expression and purification of ESAT-6

Expression of ESAT-6 was performed according to a protocol similar to that used for CFP-10 (outlined above), with some modifications. Similarly, primary cultures were set up form glycerol stocks made by Lia Rotherham (Rotherham et al., 2012). The stocks consisted of *E. coli* BL21 (DE3) cells (Novagen) transformed with the T7 promoter-based expression plasmids pMRLB7, expressing full-length N-terminal hexahistidine tagged ESAT-6 protein.

2.3.2.1 Expression of recombinant ESAT-6

The ESAT-6 overnight starter cultures were set up as described for CFP-10 above. Expression of ESAT-6 was induced at mid-log phase (OD = 0.6, at 600 nm), by the addition of IPTG (Sigma) at a final concentration of 0.4 mM. However, unlike the CFP-10 cultures, the induced ESAT-6 cultures were incubated overnight at 25°C. The post-induction incubation at 25°C was conducted to increase the solubility of ESAT-6. ESAT-6 enters inclusion bodies when induced at 37°C (Renshaw et al., 2002). Previous studies about expression of MTB antigens reported that reducing the culture growth temperature helps enhance the solubility of the expressed protein (Chaudhary et al., 2005, Kulshrestha et al., 2005). Following the overnight incubation, harvesting of the cells was achieved by centrifugation at 6500 rpm for 20 minutes at 4°C. The ESAT-6 -containing cell pellets were stored overnight at -80°C prior to the purification step.

2.3.2.2 Purification of recombinant ESAT-6

ESAT-6 was purified from the supernatant using a protocol similar to that used for CFP-10, except that the Ni-NTA column was washed with 20 mL of Buffer B containing 40 mM and 50 mM imidazole, respectively, prior to elution of the ESAT-6 protein in 300 mM imidazole as described above.

2.3.3 Concentration of Proteins: CFP-10 and ESAT-6

The CFP-10 and ESAT-6 proteins were concentrated through a buffer exchange process using the Vivaspin-6, 3 KD cut-off membrane columns), according to manufacturer's instructions (Sartorius Stedim Biotech). The buffer exchange process consisted of repetitive steps of loading of the column with Hepes buffer (200 mM Hepes, 200 mM NaCl, pH 2.5), followed by centrifugation at 4°C and 4500 rpm, and subsequent decanting and loading of fresh buffer after every centrifugation step. Following six repeats of this process, the protein was finally displaced from the membrane by gently pipetting the 1mL buffer remaining in the sample loading part of the column over the column membrane. The concentrated protein was collected in 1mL fractions and stored at -80°C.

2.3.4 Analysis of CFP-10 and ESAT-6 proteins by Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS–PAGE)

SDS-PAGE was used to resolve the proteins according to their molecular weight. SDS is an anionic detergent known to denature secondary and non–disulfide–linked tertiary structures, and confer a negative charge to each protein in proportion to its mass. A discontinuous buffer system with a stacking and resolving gel was used to increase the resolution of protein separation during electrophoresis. The gels were prepared using the Bio-Rad mini-Protean 3 system. A resolving gel was prepared first with the composition described in Table 2-1. The gel was poured into the gel cast, and subsequently layered with 100% ethanol to ensure an even surface. Once the resolving gel had set, the ethanol

was removed and a stacking gel (Table 2-2), was layered over the resolving gel and allowed to set.

Table 2-1: Composition of Resolving gel for SDS-PAGE

Separating gel solution (17%) 10 mL	
1.5 M Tris–HCl [Sigma], pH 8.8 (0. 1% SDS) [Sigma]	2.5 ml
Acrylamide/Bis–acrylamide (40% w/v) [Sigma]	4.25 ml
ddH_2O	3.14 ml
10% (w/v) ammonium persulfate (APS) [Sigma]	100 µl
Tetramethylethylenediamine (TEMED) [Thermoscientific]	10 µl

Table 2-2: Composition of Stacking gel for SDS-PAGE

Stacking gel solution (4%) 5 mL	
0.5 M Tris–HCl (Sigma), pH 6.8, (0.1%) SDS (Sigma)	1 ml
Acrylamide/Bis–acrylamide (40% w/v)	0.5 ml
ddH_2O	3.45 ml
10% (w/v) ammonium per sulfate (APS)	50 µl
TEMED [Thermoscientific]	5 µl

Once the gels had set, each protein sample was mixed a 1:1 ratio with a 2 × Laemmli samples buffer (63 mM Tris-HCl (pH 6.8, 10% Glycerol, 2% SDS, 2% (v/v) 2-mercaptoethanol, 0.1% (w/v) Bromophenol Blue) prior to denaturation at 95°C for 5 minutes. Following denaturation, the protein samples were resolved on the discontinuous buffer gels using the mini-Protean 3 system (Bio-Rad). The gels were

electrophoresed in 1×Tris-Glycine buffer (25 mM Tris base, 192 mM Glycine, pH 8.3) at 30mA for 1 hour. The leading ions of the gel (as indicated by the blue dye) were allowed to diffuse through the gel before electrophoresis was stopped. The gels were then removed from the Bio-Rad mini Protean system, and rinsed in distilled water before staining with Coomassie brilliant blue R-250 (0.25% Coomassie Blue R-250 (Sigma), 50% Methanol (Merck) and 10% acetic acid (Merck). Following sufficient staining, the gels were destained using a solution of 50% Methanol (Merck), 40% acetic acid (Merck), and 10% ddH_2O and viewed on the Molecular Imager Chemidoc XRS+ Imaging system (Bio-Rad).

2.3.5 Western blot analysis of CFP-10 and ESAT-6 proteins

Western blot analysis was used to confirm the presence of CFP-10 and ESAT-6 using their cognate antibodies. Purified CFP-10 and ESAT-6 were respectively resolved on a 17% SDS-PAGE at 30 mA as described above. The proteins were then electro-blotted onto a Hybond™-enhanced chemi-luminescence (ECL) nitrocellulose membranes (GE Healthcare) using a Trans-Blot® Semi-Dry (SD) Transfer Cell (Bio-Rad). To limit non-specific binding, the membranes were blocked by incubation in PBS supplemented with 0.005 % (v/v) Tween-20 (Sigma, USA) and 5% (w/v) fat-free milk powder for 1 hour at room temperature. Blocking was followed by a 1 hour incubation at room temperature, with a 1:15,000 dilution of the respective primary (anti-CFP10 polyclonal antibody (provided by Megan Lucas, Colorado State University) or a monoclonal mouse anti-ESAT-6 antibody (Santa Cruz Biotechnology). The membranes were washed three times with wash buffer (PBS containing 0.005% Tween-20), and incubated for 1 hour with 1:15,000 dilution of the horseradish peroxidase-(HRP) conjugated secondary antibodies: HRP goat anti-rabbit IgG and HRP goat anti-mouse IgG, respectively (Santa Cruz Biotechnology). Following three washes, the proteins were detected with ECL Advance™ Western Blotting Detection kit (GE Healthcare) according to the manufacturer's instructions. Western blot imaging and analysis was conducted on the Bio-Rad gel doc system, using the Image J software.

2.3.6 Determination of protein concentration

The Pierce® bicinchoninic acid (BCA) Protein Assay Kit was used to quantify ESAT-6 and CFP-10, respectively. The BCA method is based on the combined reduction of Cu+2 to Cu+1 by protein in an alkaline medium (the biuret reaction), and the highly sensitive and selective colorimetric detection of Cu+1. The chelation of two molecules of BCA with one cuprous ion in the assay results in a purple-coloured reaction product. This water-soluble complex exhibits a strong and nearly linear absorbance at 562 nm over a broad working range (20-2,000 μg/mL). The BCA protocol in the current study was carried out according to the manufacturer's manual. Briefly, a standard curve for the protein micro assay was set up according to OD595 nm against a concentration of Bovine Serum Albumin (BSA) standards. Following incubation at 37°C for 30 min, absorbance was measured against the reagent blank and the unknowns (ESAT-6 and CFP-10) were extrapolated from the standard curve.

2.3.7 *In vitro* formation of the CFP-10 and ESAT-6 heterodimer complex

The purified ESAT-6 and CFP-10 recombinant proteins were mixed in 1:1 molar amounts (5 μM) in a phosphate buffer (25 mM NaH_2PO_4, 100 mM NaCl, pH 7.5), as previously described by (Renshaw et al., 2002). The ESAT-6.CFP-10 dimer was then analysed by native or non-denaturing PAGE. The non-denaturing gels and running buffer were prepared in the same way as the SDS-PAGE, but without the SDS, which is the denaturant.

2.4 DNA Technology

The full length aptamers used in this study were generated from a previous study (Rotherham et al., 2012) through the SELEX process.

2.4.1 Isolation of Plasmid DNA

To produce template DNA plasmids for the intended production of dsDNA, glycerol stocks of the aptamer clones were used as starting material for the starter cultures. The

resultant aptamer pools had been ligated into plasmid vectors (pGEM-T Easy vector) (Promega) to produce individual aptamer clones (Rotherham et al., 2012). Starter cultures for each aptamer were prepared by inoculating a single colony from an aptamer glycerol stock into 10ml of LB medium (Sigma) supplemented with 100 µg/ml ampicillin (Sigma). The starter cultures were grown overnight (~12 hours) in a shaking incubator, at 37°C with vigorous shaking (300 rpm. Each starter culture was inoculated into 200mL of LB medium, and grown at 37°C with vigorous shaking (300 rpm). The cultures were grown in flasks with a volume capacity of at least four times the volume of each culture to allow for proper aeration. The bacterial cells were harvested by centrifugation at 6000 × g for 15 minutes at 4°C.

Cell lysis and plasmid preparations were carried out using the Plasmid Maxi Kit (QIAGEN) according to the manufacturer's protocol. The concentration of the plasmid DNA was quantified using the NanoDrop® 1000 spectrophotometer. The purity of the plasmid DNA was assessed by 1% agarose gel electrophoresis with Ethidium Bromide staining.

2.4.2 PCR amplification of Plasmid DNA

Polymerase chain reaction (PCR) was used to produce and amplify the dsDNA required to make functional ssDNA for further characterisation of the aptamers. The template DNA used in this experiment was from the plasmids described above. The following primers; forward primer: 5′-biotin-GCCTGTTGTGAGCCTCCTAAC-3′, reverse primer: 5′-phosphate-GGGAGACAAGAATAAGCATG-3′, were purchased from Integrated DNA Technologies (IDT). Usually, about 100ng of template DNA was used per PCR reaction. All PCR reactions were set up in 100 µl of reaction volume (Table 2-3) The PCR protocol was as follows: initial denaturing step at 95°C for 2 minutes; twenty cycles of denaturing step at 95°C for 30 seconds, an elongation step at 56°C for 30 seconds and an extension step at 72°C for 30 seconds; final extension was performed at 72°C for 2 minutes.

Table 2-3: Composition of one PCR reaction mix (of 100 ml)

Stock	Final concentration	Volume
Template DNA	10–100 ng	0.5–1 µl
5×Taq Buffer	1×	20 µl
MgCl2	1.5mM	14 µl
10 mM GeneAmp® dNTPs (Invitrogen)	0.2 mM	2 µl
100 µM forward primer	1 µM	1 µl
100 µM reverse primer	1 µM	1 µl
Taq Polymerase (5U/µl)	25 U/ml	0.5 µl
H_2O		ad 100 µl

2.4.3 Purification of PCR product

To purify the PCR product and remove free nucleotides from the DNA samples, a Promega PCR clean-up kit was used according to the manufacture's protocol. The concentration of the dsDNA was calculated from absorbance readings at 260 nM. The purified dsDNA was analysed by non-denaturing PAGE.

2.4.4 Lambda exonuclease digestion for generation of ssDNA

To generate ssDNA from the dsDNA, the PCR product was digested using lambda (λ) exonuclease, as previously described (Avci-Adali et al., 2009). The λ exonuclease is a 5′ to 3′ exodeoxyribonuclease; an enzyme that selectively digests the phospohorylated strand of dsDNA (Figure 2-1). The enzyme is known to exhibit greatly reduced activity on ssDNA and non–phospohorylated dsDNA (Higuchi and Ochman, 1989). In the current study, 10 U of λ exonuclease (New England Biolabs) were used to digest 6.6µg of purified dsDNA in a total volume of 100 µl with 1× λ exonuclease buffer. The exonuclease mixture was incubated at 37°C for 4 hours, after which the reaction was

terminated by incubation at 70°C for 10 min. The resultant ssDNA (biotinylated strand) was purified using the Nucleospin® Extract II kit (Machery-Nagel), and analysed by non-denaturing polyacrylamide gel electrophoresis.

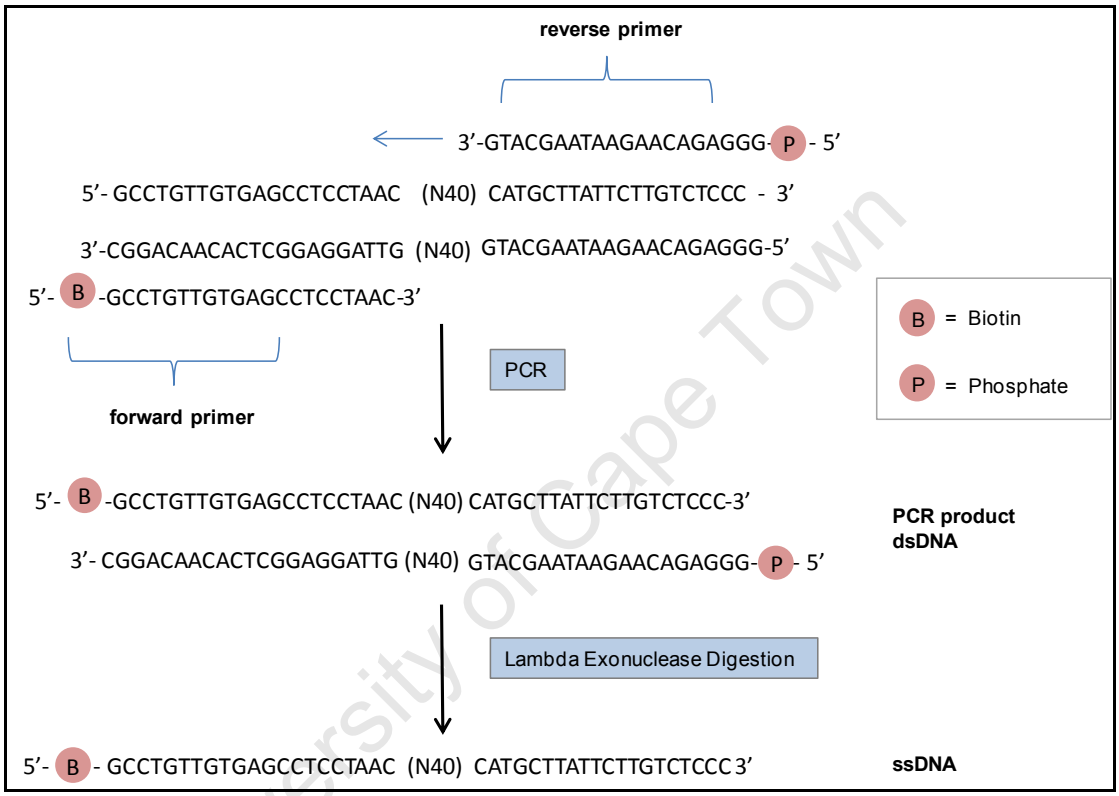


Figure 2-1: Schematic illustration of ssDNA production of selected anti-ESAT-6.CFP-10 aptamers. The dsDNA was produced by PCR using the relevant primers. The reverse primer, which was 5′ phosphate-modified enables Lambda exonuclease digestion, which selectively digests the phosphorylated strand of the dsDNA. The remaining strand contains a primer with a 5′ biotin modification to enable downstream aptamer binding affinity assays.

2.4.5 Analysis of DNA by gel electrophoresis

The DNA aptamers were analysed by electrophoresis. Agarose gels were used to resolve the plasmid DNA, while polyacrylamide gels were used for the PCR product and the ssDNA.

2.4.5.1 Agarose gel electrophoresis

Agarose gel electrophoresis is commonly used to resolve larger fragments of DNA. Usually, 2% agarose gels are used for dsDNA with less than 100 bp, and 1% gels for larger DNA fragments. The plasmids of the aptamer clones in the current study were resolved on 1% agarose gels. To prepare the gel, agarose powder (Sigma) was mixed with 1× TBE (89 mM Tris-borate, 2 mM EDTA, pH 8.3) to the desired concentration. For 1% gels, 1g of agarose powder was dissolved in 100 mL of 1× TBE, and heated in a microwave until completely melted. The gel mixture was left to cool to about 60°C, after which Ethidium Bromide was added to the agarose mix to a final concentration of 0.5µg/ml.

To prepare the DNA samples for loading on the gel, 6× DNA loading buffer (Fermentas) was added to each sample, and 50 ng of DNA were loaded in each well. The MassRuler DNA ladder mix (Thermo Scientific) was loaded in a separate well and used as a reference with which to confirm the molecular weight of the DNA samples. The gel was electrophoresed in 1× TBE (89 mM Tris-borate, 2 mM EDTA, pH 8.3) buffer at 100V. The gels were visualized under ultraviolet (UV) light using the Molecular Imager Chemidoc XRS+ Imaging system (Bio-Rad).

2.4.5.2 Non-denaturing (Native) polyacrylamide gel electrophoresis

Polyacrylamide gels are generally used to separate shorter nucleic acids. In the case of non-denaturing or native polyacrylamide gels, nucleic acids migrate at different rates, based on their secondary structure. In contrast, nucleic acids run as single-stranded fragments on denaturing gels. Under denaturing conditions, the secondary structure of the nucleic acid does not form, and thus only the length of the molecule will affect mobility.

The nondenaturing polyacrylamide gel mixtures used in this study were prepared using the compositions outlined in Table 2-4. The DNA samples were prepared for electrophoresis by adding the 6× DNA loading buffer (Fermentas) to each sample.

About 50 ng of DNA were loaded in each well, and a GeneRuler Low Range DNA Ladder (Fermentas) was loaded in a separate well to confirm the molecular weight of the DNA samples. The gels were electrophoresed in 1× TBE buffer at 100V followed by Ethidium Bromide staining and visualization using Molecular Imager Chemidoc XRS+ Imaging system (Bio-Rad).

Table 2-4: Polyacrylamide gel composition for DNA analysis

10% Non-denaturing Polyacrylamide gel (12%) 10 ml	
Acrylamide/Bis–acrylamide (40% w/v) [Sigma]	3 ml
10 × TBE buffer (1 × final concentration)	1 ml
10% (w/v) ammonium per sulfate (APS)	100 µl
TEMED	10 µl
ddH_2O	4.9 ml

2.5 Characterisation of molecular interactions of full length aptamers with target proteins

Six of the aptamers that showed the best binding profiles during the initial screening for binding to recombinant target using ELONA (Rotherham et al., 2012) were selected for further characterisation. Given that the yield of ssDNA from lambda exonuclease digestions was insufficient for downstream assays, chemically synthesised aptamers were obtained from IDT. The aptamers were synthesised with a 5′-Biotin modification to enable for detection using Streptavidin-conjugated HRP-catalysed reactions in ELONAs.

2.5.1 Surface Plasmon Resonance (SPR) technology-based determination of aptamer dissociation constant (K_D) values

To determine binding affinities of the aptamers to the target, SPR technology was used. Briefly, target protein (CFP-10) was immobilised unto a CMD 500 biosensor chips (Xantec) on a BIAcore® 3000 biosensor instrument (GE Healthcare). This was followed by the injection of serial dilutions of the analyte (respective aptamers) over the ligand surface at a constant flow rate. The ligand-target interactions were then measured by the SPR instrument as changes in the refractive index, recorded as response units (RU).

2.5.1.1 Preparation of chip surface and immobilisation of ligand

To activate the carboxy-methylated dextran surface of the biosensor chips, 50 µl of a 1:1 combination of 0.5 M N-hydroxysuccinimide (NHS), 0.2 M N-ethyl-N'-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) (Biacore, GE Healthcare) was injected at a flow rate of 10 µl/minute, over all the four flow cells of the biosensor chip. The activation step allows for amine coupling of the protein to the surface of the chip. Activation was followed by injection of 50 µl of the in-house-purified CFP-10 monomer diluted to 0.015 mg/ml/ (1.5 µM) in 100 µl acetate buffer (10 mM, pH 4.5), into three of the four flow cells (1 to 3). Approximately 3000 RU of CFP-10 were coupled. Flow cell 4 was used as the control (reference flow cell with no protein). Ethanolamine-HCl (1 M, pH 8.0) was injected over the immobilised protein to block the remaining activated carboxymethyl groups. This was followed by the injection of 35 µl of 10 mM NaOH (regeneration solution) to reduce any non-specific binding that remain after the ethanolamine-HCl blocking step.

2.5.1.2 Preparation of aptamers

Aptamer stocks were prepared by reconstituting the lyophilised chemically synthesised ssDNA aptamers (IDT) in water to a final concentration of 100 µM. The ssDNA aptamers were run on a native polyacrylamide gel for independent validation and verification of size and purity. The ssDNA aptamers were prepared for SPR experiments

in refolding (selection) buffer (1× final concentration); HMCKN buffer consisting of 10 mM Hepes pH 7.4, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 2.7 mM KCl, 150 mM NaCl). The aptamers were diluted to give a series of concentrations: 31 nM, 62 nM, 125 nM, 250 nM, and 500 nM.

2.5.1.3 Injection of aptamers over the CFP-10-coated sensor chip surface

Prior to injecting the aptamers, 50 µl of 1× HMCKN buffer was injected across all four flow cells; to control for the buffering effect. The aptamers were heated to 95°C for 5min, and rapidly cooled at 4°C, briefly settled at room temperature before they were injected over the surface at a flow rate of 10 µl/min (to minimize mass transfer limitations). ssDNA aptamers at concentrations 500, 250, 125, 62, 31 nM, were then injected over all four flow cells. The binding events (aptamer-protein interactions) were recorded by the SPR instrument. The aptamers were allowed a 600 seconds dissociation time before the sensor chip was regenerated using 10µl of 10mM NaOH after each aptamer binding step. This was done for each of the six selected ssDNA aptamers. Bulk shift and non-specific interactions with the CFP-10 were subtracted using the response from a reference flow cell (flow cell 4), which did not contain CFP-10. Figure 2-2 illustrates the flow of events during the process of characterising aptamer-CFP-10 interactions.

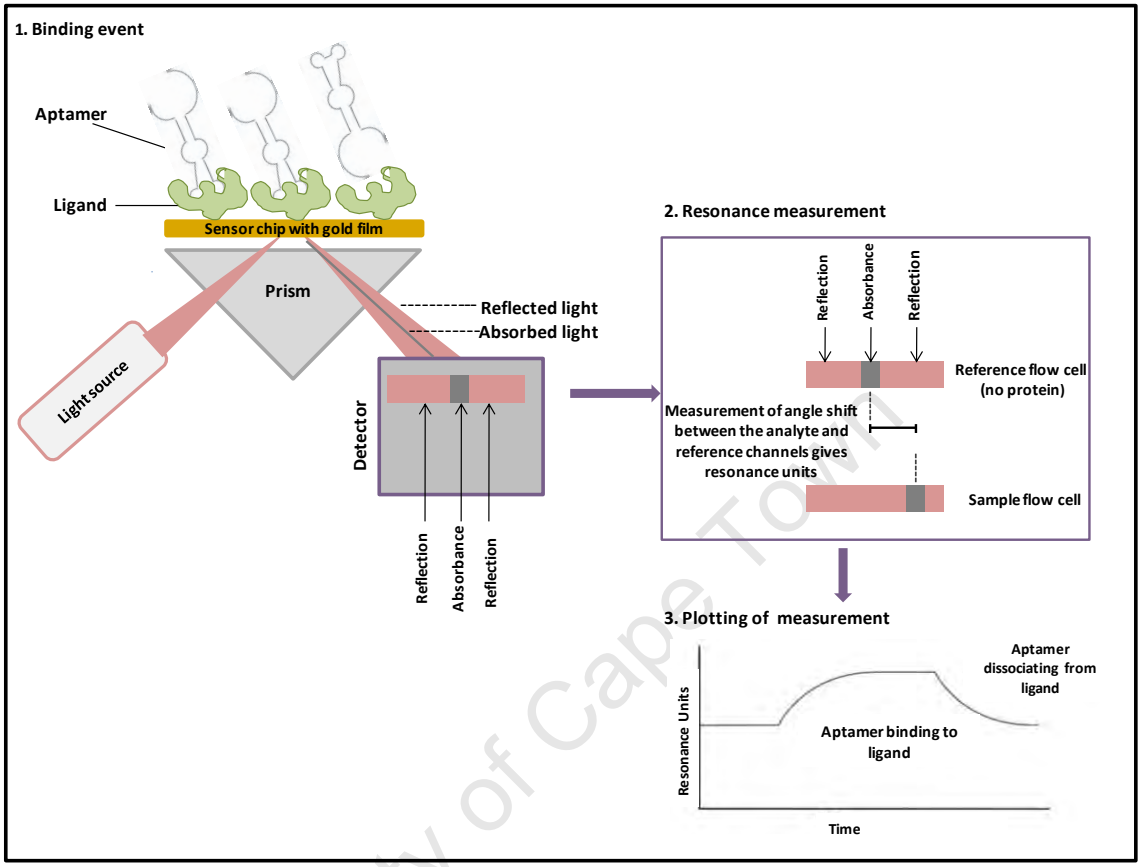


Figure 2-2: Schematic representation of the process of evaluation of ssDNA aptamer-CFP-10 interaction by SPR

Analysis of data obtained in these experiments was performed using the BiaEvaluation Software (BIAcore). The software allowed for calculations of the association and dissociation constants, from which equilibrium dissociation constant (K_D) values were determined. Resulting data were prepared for kinetic analysis by resetting the time and RU to zero, respectively, before the first injection. To control for refractive index changes and nonspecific binding, the responses from the control flow cell (no protein) were subtracted from those generated for the aptamer-CFP-10 interaction. The affinity of the aptamers, as described by the K_D, was determined by globally fitting the binding data to the kinetic simultaneous ka/kd model, assuming Langmuir (1:1) binding. The

model follows pseudo-first order kinetics and assumes that binding is equivalent and independent for all binding sites. The Chi^2 values for the dissociation constants were determined, as a measure of the reliability of data.

2.5.2 Characterisation of the ssDNA aptamers-ESAT-6/CFP-10 heterodimer interactions using EMSA

The kinetic interactions between the ssDNA aptamers and the ESAT-6.CFP-10 heterodimer were investigated by EMSA. SPR could not be used because the ESAT-6.CFP-10 dimer complex is held together by salt bridges and this poses a challenge in the stability of the protein during SPR experiments. Due to the relatively harsh pH conditions of the SPR ligand buffer, and the constant flow of buffers in the microfluidic environment of the sensor chip on which the protein is immobilised, as well as the regeneration of the sensor chip using NaOH after every aptamer binding event, the dynamics in the SPR instrument may not be favourable for the stability of the heterodimer. As an attempt to circumvent the above-mentioned technical hurdles, EMSA was considered a good alternative technique to assess the molecular interactions and stoichiometric interactions between the aptamers and the ESAT-6.CFP-10 dimer.

In a typical EMSA experiment, a nucleic acid is labelled at the 5′-end label with radioisotope ^{32}P, and subsequently incubated with increasing amounts of target protein, followed by electrophoresis of the aptamer-protein complexes, and analysis of mobility using a phosphoimager or autoradiograph (Figure 2-3). From this, the amount of bound and unbound aptamer can be quantified.

The EMSA protocol used in the current study was adapted from (Buratowski and Chodosh, 2001). The process is divided into several parts, namely, radioactive labelling of the ssDNA, preparation of non-denaturing gel, binding reaction, electrophoresis and gel analysis.

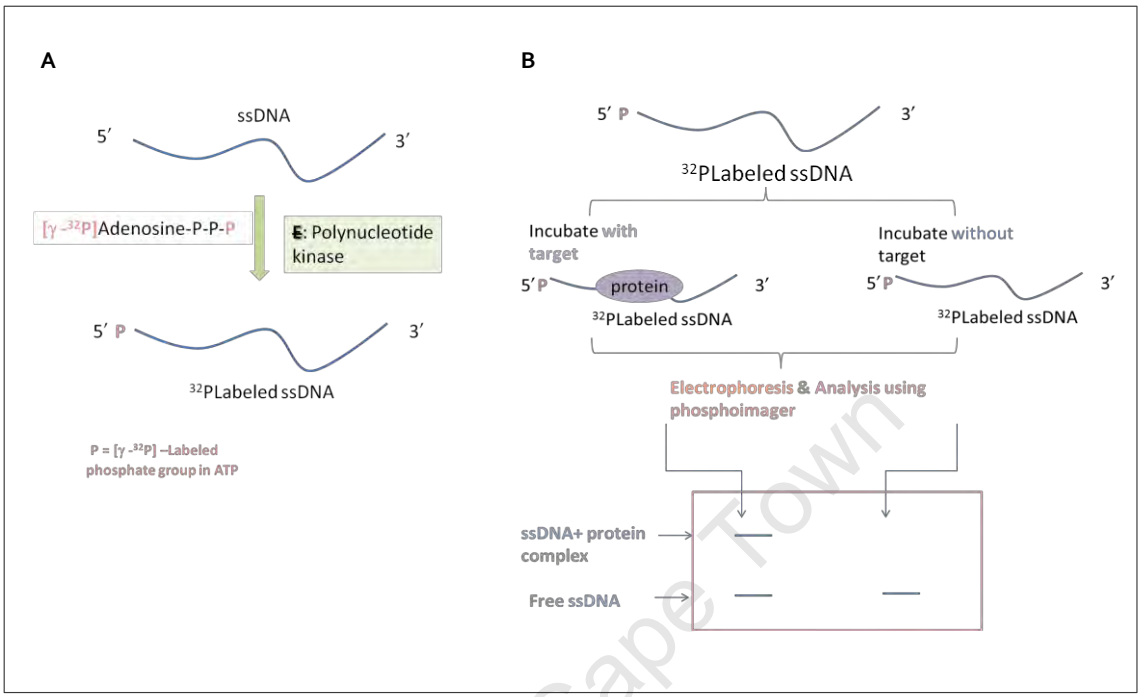


Figure 2-3: Schematic flow diagram describing the EMSA process. A ssDNA probe is radiolabelled **(A),** after which the probe is incubated with its cognate protein and subjected to gel electrophoresis **(B).** Finally, the electrophoresed samples are imaged and analysed for mobility shift.

2.5.2.1 Labelling of the ssDNA aptamer probe

The in-house produced CSIR 2.11 aptamer was used for the EMSA experiments. For the 5′ end labelling reaction, 50 pmol of CSIR 2.11 was mixed with 50 pmol (150 µCi) γ-^{32}P ATP (PerkinElmer) (specific activity 3,000 Ci/mmol), 10 × Polynucleotide Kinase (PNK) Buffer, 20 U of the enzyme PNK (Perkin Elmer), in a 1.5 ml Eppendorf tube, and nuclease-free water was added to a final volume of 20 µl. The mixture was incubated for 60 min at 37^{o}C, after which the reaction was stopped by adding 1 µl of 0.5 M EDTA. The labelled DNA was separated from free ATP using Ambion nuclease-free NucAway Spin Columns (Invitrogen), according to the manufacturer's instructions. To determine if the labelling was effective, the labelled DNA was electrophoresed on a 8% non-denaturing polyacrylamide gel, which had been pre-run in 1 × running buffer, for 60

min at 100 V. Electrophoresis of the labelled CSIR 2.11 was carried out at 30 mA for 30 to 45 min.

2.5.2.2 Preparation of binding reactions

The binding reactions were prepared in 1.5 ml Eppendorf tubes. Radio-labelled CSIR 2.11 was mixed with 1 × HMCKN buffer to a final concentration of 3 ng/µl. To allow for refolding, the aptamer was heat-denatured at 95^{o}C for 5 minutes, and quickly cooled on ice for 5 minutes, and finally incubated at room temperature for 5 minutes prior to addition of protein. Increasing concentrations of protein were then added to the mix. The first aptamer-dimer reaction attempts were carried out with a lower concentration range: 25, 50, 100, 150, 200, 250, 300, and 350 nM. Subsequent attempts were conducted using higher concentration ranges 250 nM, 350 nM, and 1 µM, and later 1, 5 and 10 µM. Glycerol was added to a final concentration of 10 % v/v, to facilitate sample loading and to enable the samples to sink and settle to the bottom of the wells of the gel. The final reaction volume was adjusted to 20 µl. The reaction mixtures were mixed gently by tapping the bottom of the tube with a finger, and incubated for 1 hour at room temperature.

2.5.2.3 PAGE analysis of aptamer-protein complexes

Following the 1 hour incubation, binding reactions were loaded (without loading buffer or dyes) into the appropriate wells of the native PAGE gel that had been pre-run for 30 minutes. A small volume of 6× DNA loading dye (Fermentas) was loaded into a separate well, and was used to monitor the progress of the electrophoresis.

2.5.2.3.1 Electrophoresis of aptamer-protein complexes

The binding reactions were electrophoresed at 30mA for 45 minutes at room temperature. The recommended minimum time required to give good separation of free probe (ssDNA) and the DNA-protein complexes is at least 30 minutes (Hellman and

Fried, 2007). Electrophoresis was then stopped before the bromophenol blue of the loading dye approached the bottom of the gel.

2.5.2.3.2 Processing and analysis of the gel

The glass plates were removed from the gel box, and carefully separated. The gel remained attached to one of the plates. The plate with the gel was laid on the bench with the gel facing up, and subsequently sandwiched with a sheet of Whatman 3 μm filter paper cut to size. The filter paper was then peeled off from the plate with the gel attached to it, and covered with a single layer of plastic wrap (without introducing any bubbles) and placed on a gel drier. The gel was then dried under vacuum for 75min: the first 15 minutes only the vacuum was applied, and the 60^{o}C heat was only introduced after the 15 minutes. Following thorough drying, the gel was placed in a cassette (Bio-Rad) and exposed, overnight, to a phosphor-imaging screen (Bio-Rad). The gel image was obtained using a phosphoimager; the Personal FX Molecular imager (Bio-Rad), at a high scan resolution of 50 microns.

2.6 Structural characterisation of full length aptamers

The mfold program (http://mfold.bioinfo.rpi.edu/cgi-bin/dna-form1.cgi) was used at default settings, and allowed for estimations of the MFE secondary structures of aptamers. The temperature was set to 37°C; salt conditions were at 1.0M NaCl. The difference in free energy between the MFE and the suboptimal structures within the same energy range (percent suboptimality), was set to 5. The factor that determines how different the secondary structures should be relative to one another (upper bound) was set to 50, while the maximum distance between base pairs was set to “no limit”.

2.7 Optimisation of full length aptamers by rational (secondary structure-guided) truncation

Two aptamers, namely CSIR 2.11 and CSIR 2.19, were selected for optimisation through truncation. The two aptamers were primarily selected for their relatively low

dissociation constants (Table 3-1). In addition, CSIR 2.11 had already been tested for its ability to accurately detect the target proteins in clinical samples (Rotherham et al., 2012). The two aptamers were truncated through a series of 5′ and 3′-end base deletions to the shortest possible sequences that retained the mfold predicted secondary structures, and potentially the functional motifs of the original full length sequences.

2.8 Computational analysis of folding behaviour of the full length aptamers versus the truncated aptamers: energetics and folding kinetics

The Vienna RNA Package algorithms were used to probe the folding dynamics and kinetics of the truncated and parental full length aptamers as previously described (Flamm et al., 2002

, Gruber et al., 2008, Lorenz et al., 2011). Specifically, the *Barriers* program (http://rna.tbi.univie.ac.at/cgi-bin/barriers.cgi) was utilised for conformational energy landscape analysis of the two ssDNA aptamers, CSIR 2.11 and 2.19 and their respective truncated derivatives.

2.8.1 Conformational energy landscape analysis of aptamers

Barrier trees were constructed using the default settings of the program, as follows: 1) basic options accounting for maximal number 50 of local minima, considering only minima with barrier height higher than 0.1 (and avoiding isolated pairs), and 2) advanced options were set for DNA energy parameters at 37°C. The program generated barrier trees representing only the 50 lowest local minima as a representative ensemble of conformations. The Barriers program uses a "flooding algorithm" that starts from an energy sorted list of all low energy conformations. This list may contain either all possible conformations or only those below a given energy threshold. From this, thermodynamic neighbours of the MFE structure are identified in the form of local minima and saddle points, which are presented on an energy barrier tree. Ultimately, the barrier tree sums up the key features of a conformational landscape which help

theoretically describe the kinetics on the landscape without the need for complex computational simulations.

2.8.2 Folding dynamics analysis: computational prediction of folding transition rates

The program *treekin* (part of *Barriers*) was used to compute the folding dynamics of the investigated aptamers. The program uses macrostates and transition rates computed by the barriers program to give a visual image of the folding process and the occupancy of meta-stable states over time (Gruber et al., 2008, Lorenz et al., 2011). To simulate the folding kinetics in the current study, the first step was to compute an energy landscape (represented in the form of a barrier tree) using the *barriers* program. This was followed by choosing an initial structure (representing the open chain), as well as a start and end time (these were user-determined) for the folding kinetics simulations using the *treekin* part of the barriers program. The *treekin* program then computed population densities as a function of time (with arbitrary units). For example, if the selected start structure (p0) is set to be macrostate 2 (the second lowest local minimum in the barrier tree), and the sum of the all population densities equals 1, the simulation will start with 100% of the initial population in macrostate 2. The transition rate matrix is constructed beginning at the start structure along the path to the MFE structure (macrostate 1 on the barrier tree). The output from these computations is a graph that shows percentage population densities (y-axis) over time in arbitrary units (x-axis). This information gives a theoretical model of the folding or refolding time from an open-chain structure to a stable target-compatible MFE structure of an aptamer.

2.9 Analysis of binding of truncated aptamers by ELONA

The rationally truncated aptamers were tested for binding to target proteins in an ELONA, under the same conditions as those used to test the binding of the full length aptamer as previously described (Rotherham *et al.* 2012). Briefly, 96-well high binding polystyrene microtiter plates (Costar, Corning) were coated with 50 µl of 500 ng of

CFP-10 monomer (prepared in 100 µl of 1 × $NaHCO_3$ coating buffer) and incubated at 4 °C overnight. The control wells were only coated with the 1 × $NaHCO_3$ coating buffer. Following the incubation period, the plate was decanted and wells were then washed three times with 150 µl of Phosphate Buffered Saline containing 0.05% Tween (PBS-T) buffer (Sigma), and blocked with 5% fat-free milk in PBS-T for 1 hour. This was followed by a further three wash steps with PBS-T. The biotinylated aptamers were prepared in 1× HMCKN (refolding buffer) to a final concentration of 500 nM, heated to 95°C and then cooled quickly to 4°C, before 50 µl aliquots of each aptamer were added to the appropriate wells and incubated for 2 hours at room temperature. Following aptamer incubation, the plate was washed four times with PBS-T, after which 100 µl of 1:15,000 Streptavidin-HRP conjugate (Kirkegaard & Perry Laboratories, KPL) was added to the wells, and incubated for 2 hours at 37 °C. Following the washing steps, 50 µl of Turbo- 3, 3', 5,5'-tetramethylbenzidine (TMB) substrates mixed 1:1 (Thermoscientific, Pierce) was added to each well, incubated for 15 min at 37 °C to enable detection of the biotin-labelled aptamers. The streptavidin-HRP: TMB reaction was stopped by the addition of 50 µl of H_2SO_4. The protein-bound aptamer-streptavidin complexes were quantified by absorbance measurements at 450 nm using the Multiskan-Go plate reader (Thermo Scientific). All the test and control samples were evaluated in triplicate, in two independent experiments.

2.9.1 Statistical analysis of ELONA-based evaluation of truncated aptamer binding to recombinant target proteins

Further analysis of the data was based on the blank-subtracted values of each data point. The triplicate values were averaged and their standard deviation calculated. A two-tailed Student's t-test was performed to determine statistical significance. To compare the performance of the truncated aptamers against their parent sequences, the relative binding of the respective aptamers was normalised to that of the 90-mer aptamer + CFP-10 (set at 1).

CHAPTER 3

RESULTS

3.1 Production of recombinant target proteins

3.1.1 Expression and purification of ESAT -6 and CFP-10

His-tagged recombinant CFP-10 and ESAT-6 proteins were successfully expressed, and confirmed through SDS-PAGE analysis of cell lysates from various stages of the protein expression process (Figure 3-1). The proteins were subsequently purified by nickel affinity chromatography, and analysed SDS-PAGE to confirm their integrity, size and purity (Figure 3-2). Purification was achieved, for the most part; however, one of the CFP-10 protein samples showed some contaminants on the gel. The identity of the proteins was determined by immunoblot analysis. The blot images showed reaction of the proteins with the respective antibodies (Figure 3-2). On average, 0.3 mg of CFP-10 was obtained per litre of culture. CFP-10 was found in the soluble fraction while ESAT-6 was found to be primarily expressed as an insoluble protein. The solubilisation of ESAT-6 was achieved through incubating the post-induction culture overnight at 25°C and the addition of urea as a chaotropic agent. The ESAT-6 yield was typically 0.1mg from 1 litre of culture. These yields were undesirably low and would be insufficient for use in functional assays with all the required optimisation steps. These yields were consistent through multiple expression and purification experiments. Therefore, to obtain sufficient protein, a number of repeat expressions and purification attempts would be required; all of which would be costly in terms of consumables and time. Consequently, large-scale produced, HPLC purified proteins were purchased for use in downstream aptamer functional assays

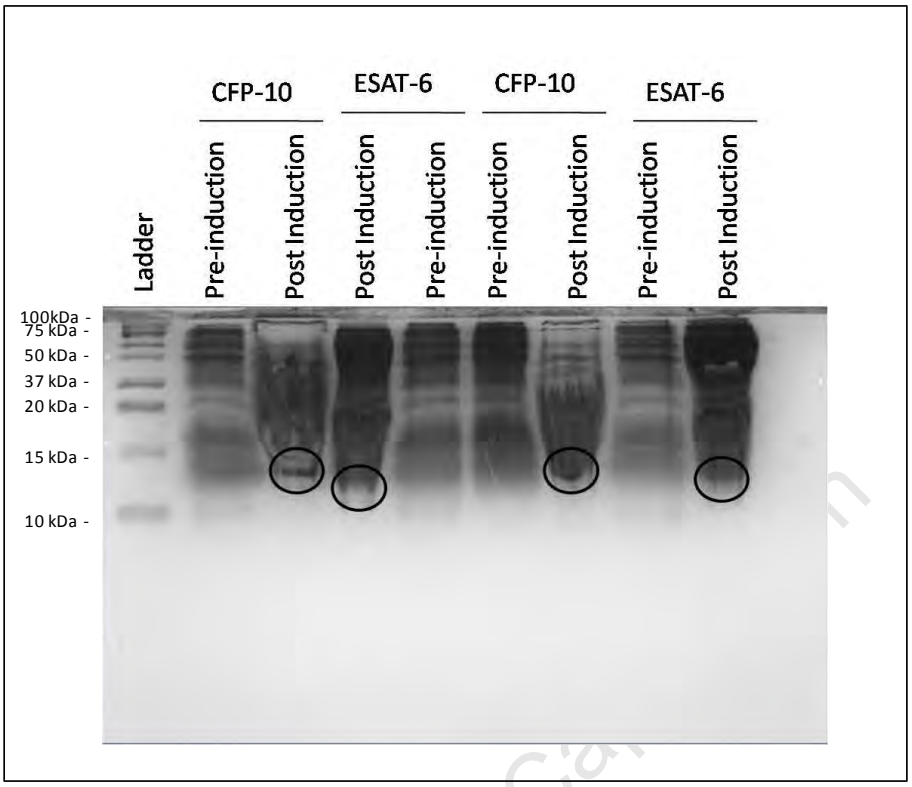


Figure 3-1: SDS-PAGE gel analysis of cell lysates from various stages of the protein expression process. Bacterial cultures of ESAT-6 and CFP-10 were induced with IPTG and the cells were analysed for protein expression. Lanes 2-9 represent the cultures sampled at different stages of expression (outlined above each well) prior to purification of protein. The circled bands represent the expressed protein of interest

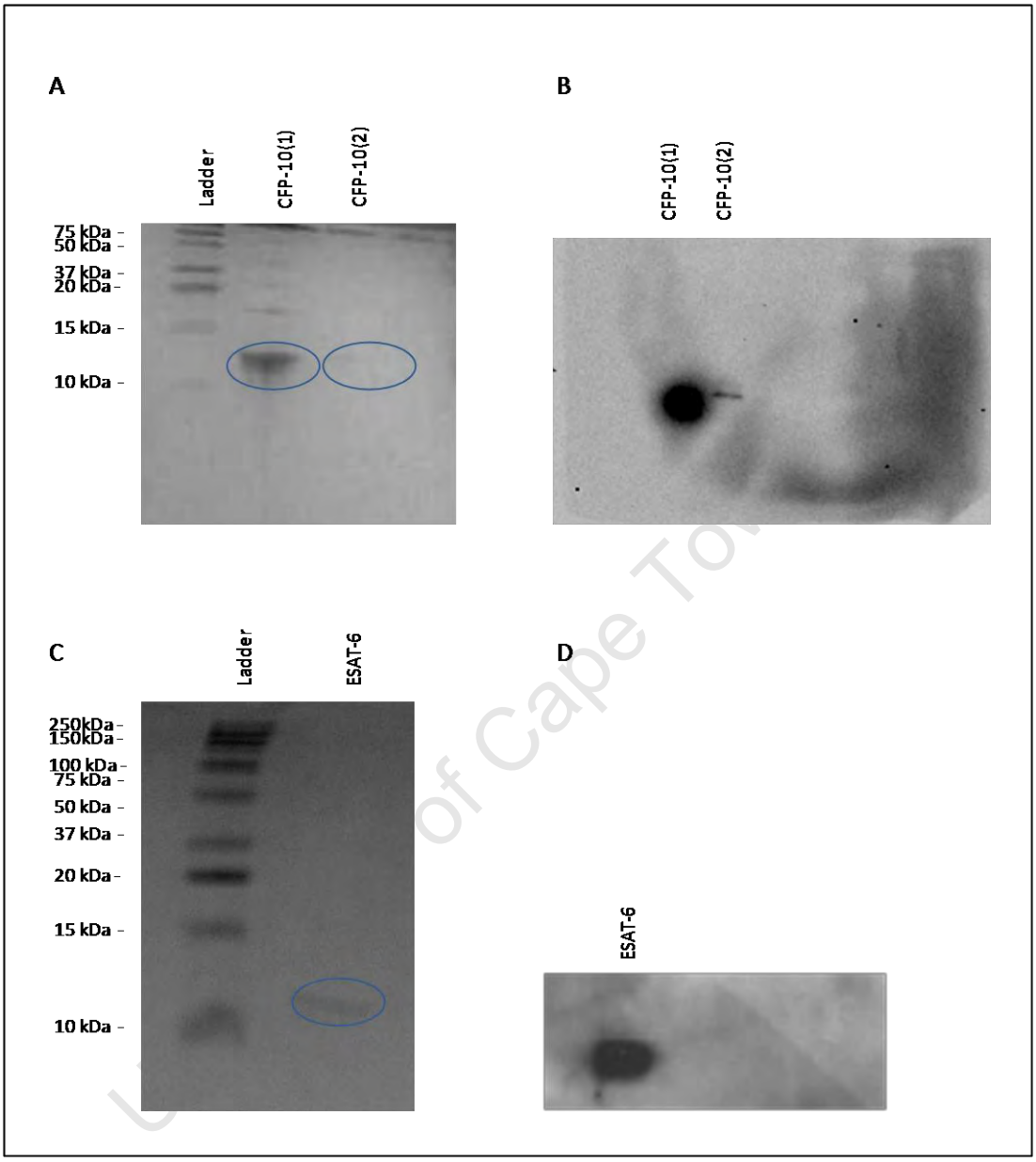


Figure 3-2: SDS-PAGE and Western Blot analysis of purified recombinant ESAT-6 and CFP-10. Bacterial cultures expressing ESAT-6 and CFP-10 were purified by IMAC, and the proteins were resolved on 17% SDS-PAGE gel. (A) SDS-PAGE of purified CFP-10. (B) Western blot analysis of CFP-10. Purified ESAT-6 is shown on SDS-PAGE (C) and the western blot (D).The bands of the purified recombinant proteins (circled) on the SDS gels, are at appropriate sizes, as indicated by the molecular weight standards.

3.1.2 In vitro production of the ESAT-6.CFP-10 heterodimer complex

To produce the ESAT-6.CFP-10 heterodimer complex in vitro, the recombinant CFP-10 and ESAT-6 monomers were mixed in equimolar amounts and dimerised in a phosphate buffer. The presence of the heterodimer complex was confirmed by non-denaturing PAGE. Two bands were observed in the heterodimer complex lane. The lower and more intense band observed between the ESAT-6 and CFP-10 monomers depicts the heterodimer complex (Figure 3-3). The higher band in the heterodimer lane may represent the CFP-10 monomer that either did not form part of the dimer complex in the first instance or had dissociated during electrophoresis. Furthermore, the ESAT-6 monomer migrated farther than both the CFP-10 monomer and the ESAT-6.CFP-10 heterodimer complex.

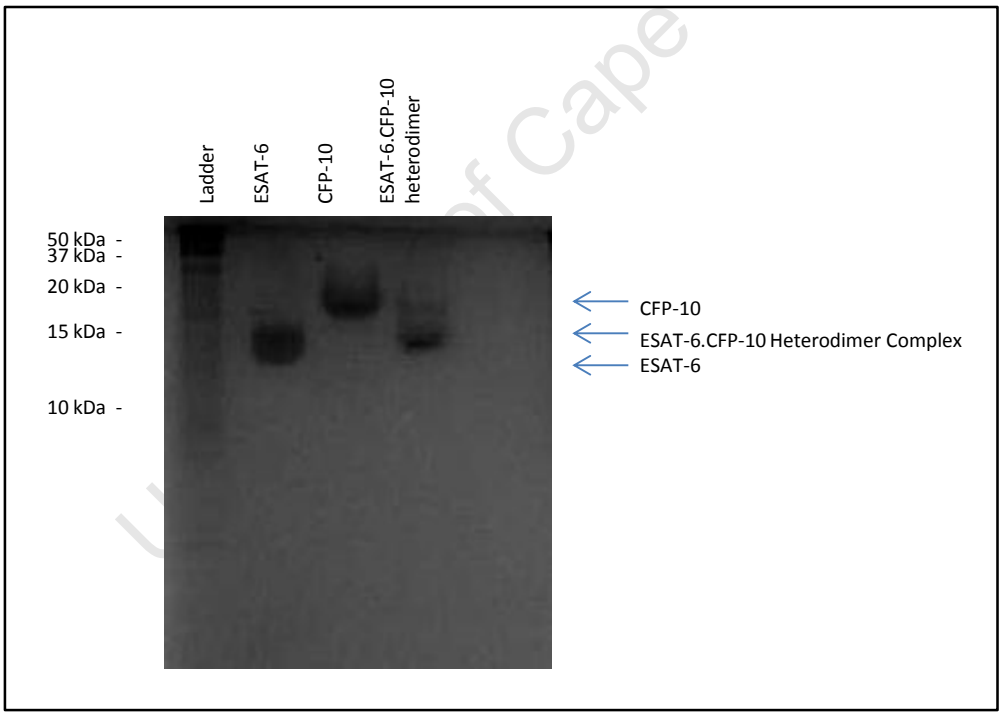


Figure 3-3: Non-denaturing PAGE analysis of the ESAT-6 .CFP-10 heterodimer. The heterodimer and respective composite monomers showed varying degrees of mobility as a result of structural conformation and charge.

3.2 Production of ssDNA of selected aptamers

As a starting point, plasmid DNA preparations were performed in order to produce template DNA for in-house production of the aptamers. The aptamer clones, which had been expressed in *E.coli* and stored as glycerol stocks were cultured, and their DNA subsequently extracted and purified. The presence and purity of the plasmid DNA of six full length aptamers were determined by agarose gel electrophoresis (Figure 3-4).

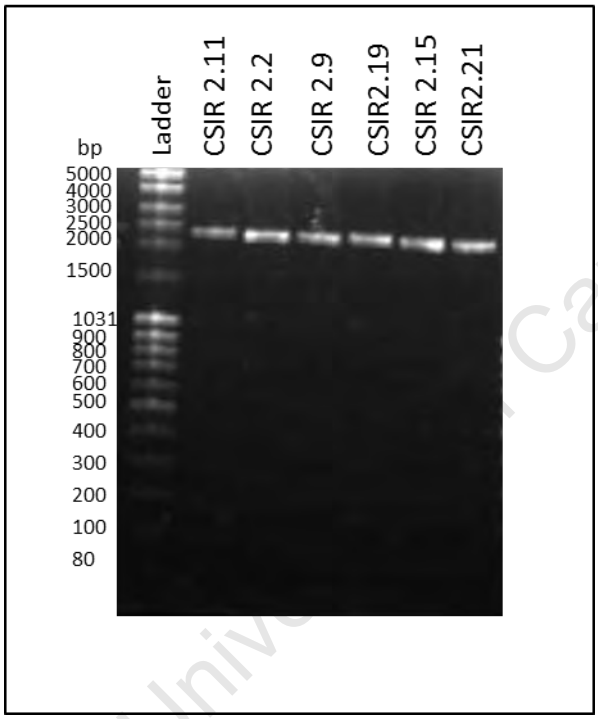


Figure 3-4: Agarose gel electrophoretic analysis of ssDNA plasmids. The isolated plasmids were purified and analysed by 1% agarose gel electrophoresis. 50 ng of DNA were loaded per well. The bands in lanes 2-7 represent purified plasmid DNA of the respective aptamers, at the expected size of around 3,105 bp.

3.2.1 PCR amplification of ssDNA aptamers

Following successful PCR amplification of selected aptamers, the quality, size and purity of the dsDNA were determined by non-denaturing PAGE (Figure 3-5). Generally, the wells showed distinct bands with no smears or laddering. The bands

in each well represent a single aptamer. The aptamers were at the expected sizes. Aptamer CSIR 2.15 has two bands which appear about 25bp apart.

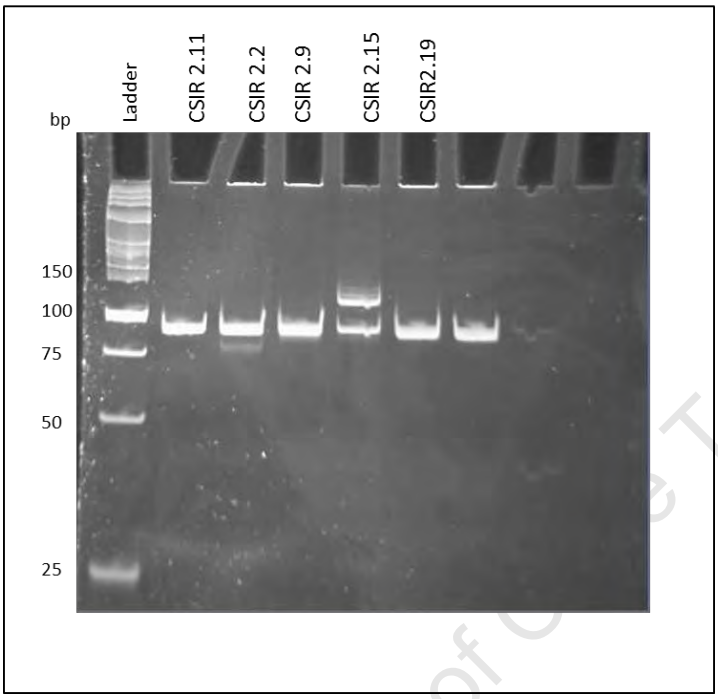


Figure 3-5: Native PAGE analysis of dsDNA after PCR amplification. The purified dsDNA was electrophoresed on a Native Polyacrylamide gel (8% Acrylamide). The bands represent the dsDNA of CSIR2.11 to CSIR 2.21. The bands of interest are at the expected size of 90bp. However, the PCR products of CSIR 2.2 and CSIR 2.15 exhibited double bands. CSIR 2.2 has a lower fainter band, while CSIR2.15 lane has one at the expected size and the other than runs slightly higher, about 25 bp apart.

3.2.2 Purification of ssDNA aptamers

To generate ssDNA, the dsDNA resulting from the PCR were digested using λ exonuclease. The enzyme selectively digested the 5'-phosphorylated strand of the dsDNA, leaving the complementary strand for use in functional assays. The efficacy of λ exonuclease-based ssDNA generation was verified by nondenaturing PAGE (Figure 3-6). The analysed samples showed no trace of dsDNA, indicating complete digestion of the phospohorylated strand. There are two bands in the CSIR 2.15 lane, as was the observed with the dsDNA form of the aptamer. The double bands suggest

the presence of two PCR products from the same plasmid; potentially resulting from one of the primers annealing to two different sequences in the plasmid.

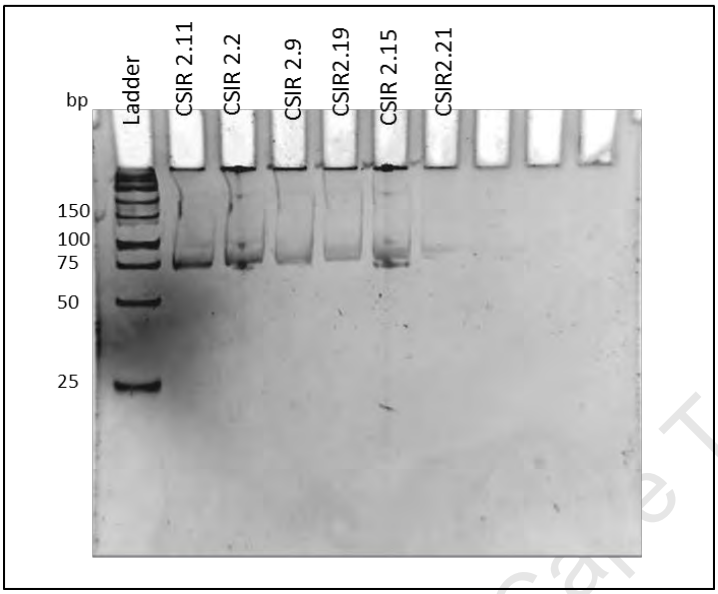


Figure 3-6: Native PAGE analysis of ssDNA after Lambda exonuclease digestion. To produce ssDNA, the dsDNA of each aptamer were treated with λ exonuclease, and subsequently purified. The purified ssDNA was then analysed by non-denaturing PAGE. The ssDNA bands run in the region of the expected sizes; between 50 and 75bp. However, given that the ladder that was used was double-stranded, this does not allow for determination of exact size of the single stranded DNA aptamers. Again, the aptamer CSIR 2.15 lane has two bands, even in the dsDNA form of the aptamer.

Although the production of ssDNA through PCR and exonuclease digestion was successful, the yields were inadequate for use in the functional assays. The maximum yields obtained in the current study were only about 40% of the maximum possible ssDNA amount that can be produced through λ exonuclease-based digestions. A similar trend was observed from several attempts of the exonuclease-based generation of ssDNA. Therefore, to circumvent the cost and time demands of producing large amounts of dsDNA via in-house PCR, and performing a number of exonuclease digestions to generate adequate ssDNA, chemically synthesised ssDNA aptamers were purchased from IDT. Only five of the six aptamers were synthesised;

CSIR 2.9 was excluded as it showed arbitrary binding kinetics (data shown in the next section). The sequences were sent to IDT, and the respective aptamers were synthesised in lyophilised form. Upon receipt, the lyophilised aptamers were reconstituted and their purity and molecular weight confirmed by non-denaturing PAGE (**Figure 3-7**). Each aptamer appeared as a single, clear band, indicating the integrity thereof.

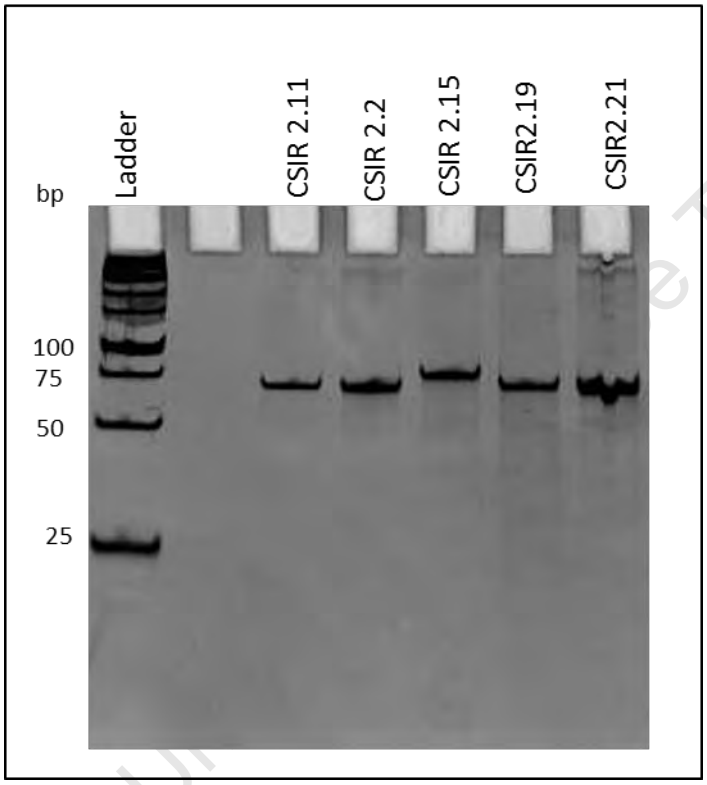


Figure 3-7: Analysis of solid-phased synthesised aptamers by non-denaturing PAGE. The bands in the labelled lanes represent each the ssDNA aptamers at the expected molecular weight. However, it is important to note that ladder was double-stranded, and therefore does not allow for determination of exact size of the single stranded DNA aptamers, but only an estimate.

3.3 Determination of binding kinetics of full length aptamers by SPR

Aptamer-CFP-10 interactions were evaluated in real time using the Biacore 3000 SPR instrument. The first step was to stably immobilise the CFP-10 protein on the

sensor chip, to allow for subsequent aptamer binding (Figure 3-8). CSIR 2.2, 2.15, 2.11 and 2.21 bound the CFP-10 target in a concentration-dependent manner (Figure 3-9). In contrast, CSIR 2.9 did not exhibit the expected concentration-dependent binding (Figure 3-10).

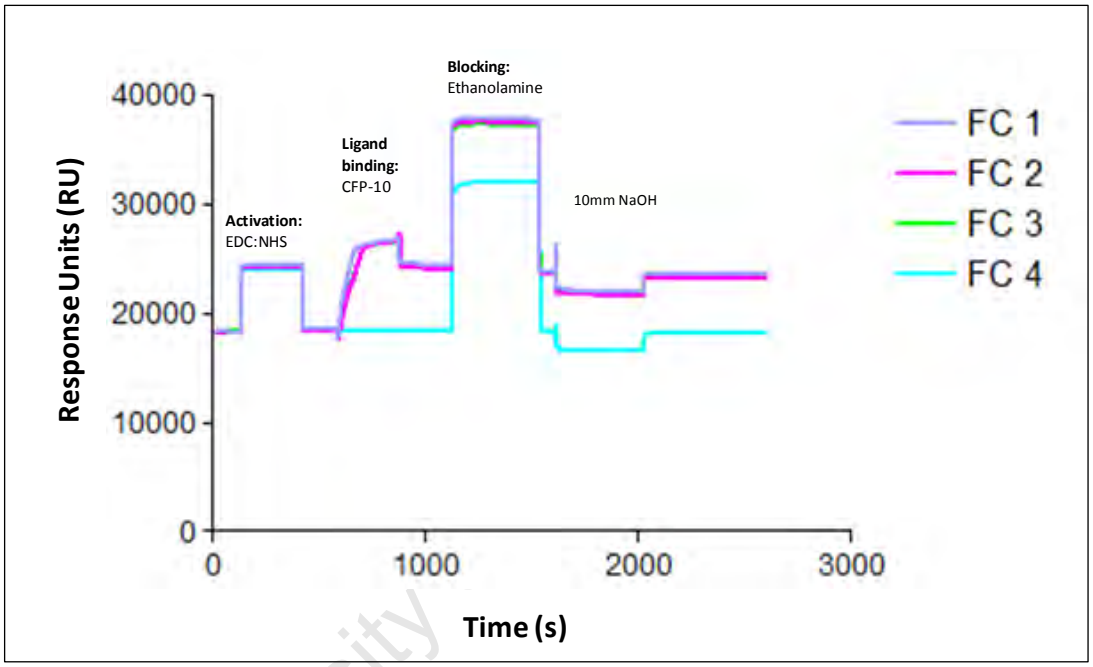


Figure 3-8: Representative Sensorgram for the coupling of CFP-10 on an SPR sensor chip surface. Activation of the sensor chip was achieved by injecting the EDC:NHS over the surface of the chip to activate the Dextran matrix for amine coupling of the ligand. The ligand, in this case, CFP-10, was amine-coupled to the chip in flow cells (FC) 1 to 3. FC4 represents the blank, "no protein" control, which was used as a reference FC from which to subtract non-specific binding. Ethanolamine was used to block remaining active sites in the FC. To wash non-covalently bound protein 10mM NaOH was used.

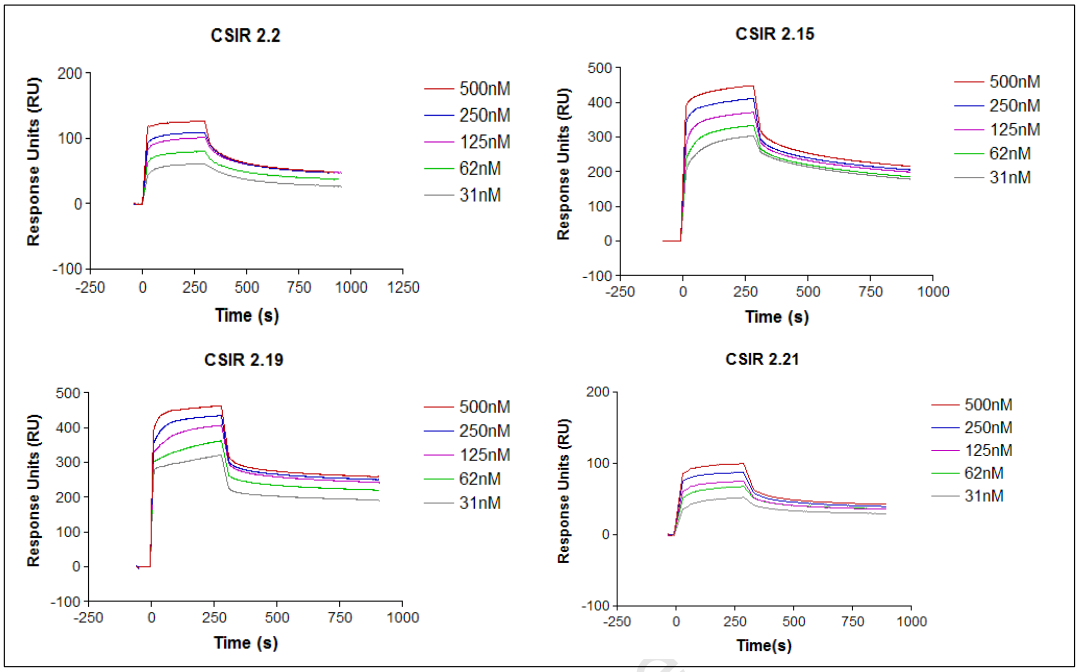


Figure 3-9: Binding Kinetics of full length aptamers evaluated by SPR. Sensorgrams show the interaction of immobilized CFP-10 on CMD500 chip (Xantec) with increasing concentrations of aptamers (31, 62, 125, 250, 250, and 500 nM). The flow rate was 10µl/min, each aptamer dilution was allowed a 600s dissociation time before regeneration of the sensor chip using 10mM NaCl. The BIAevaluation software was used to analyse the data. The blank flow cell and the buffering effect were corrected for in the analysis. The K_D values, which were calculated assuming the Langmuir 1:1 model, ranged between 1.6 and 21.5nM.

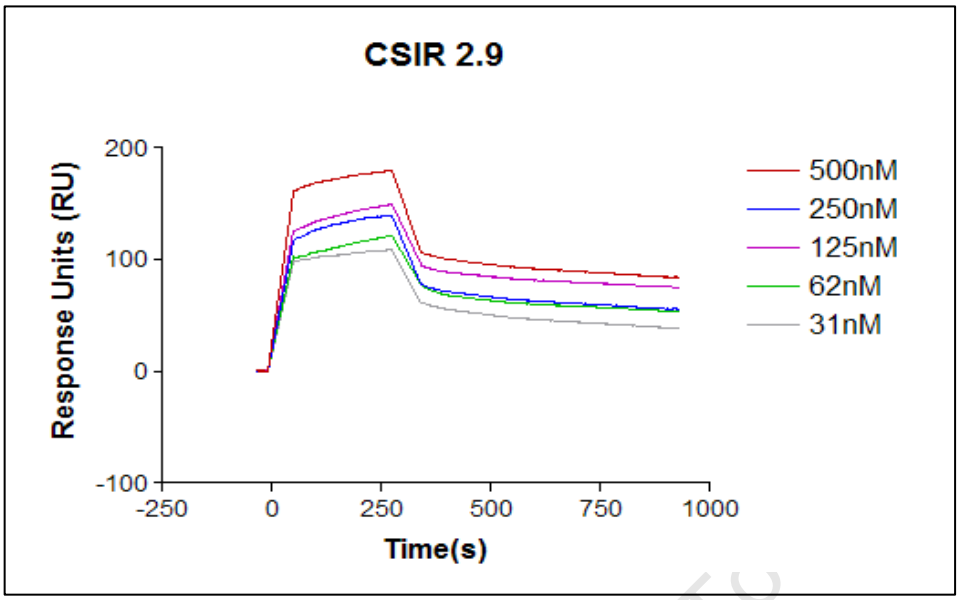


Figure 3-10: Sensorgram showing binding Kinetics of full length CSIR 2.9 aptamer evaluated by SPR. The RU at 250nM was lower than the RU at 125nM, and the analysis did not indicate concentration-dependent biding of the aptamer to the immobilised CFP-10

The kinetic parameters and associated statistics of the triplicate measurements of aptamer-CFP-10 interactions were calculated using the BIAevaluation software. The K_D values depicting the ratio of the 'on' and 'off' rates of each aptamer were globally fitted for the five different concentrations using the Langmuir model assuming 1:1 ligand: analyte interaction. Five of the six full length aptamers had K_D values within the low nanomolar range (Table 3-1), indicating tight binding of the aptamer to CFP-10. A good fit to the Langmuir model was obtained for these five aptamers, as demonstrated by the Chi^2 values (Table 3-1). Each Chi^2 value shown in the table is the highest of the three values for each data set (from the three flow cells, Appendix A 1.1.). The Chi^2 is a statistical measure of closeness of fit of experimental data to the theoretical model used to determine the K_D, i.e. it describes how well results agree with curve-fitting of the model used. Chi^2 values lower than the noise RU (in this case, ~ 20 RU), indicate a good fit. Conversely, the data obtained from the kinetic analysis of CSIR 2.9 did not fit the 1:1 Langmuir model.

Thus a K_D value for this aptamer could not be derived under the same conditions as those of the other five aptamers.

Table 3-1: Dissociation constant (K_D) values of the full length aptamers obtained through Biacore analysis with relevant statistics

Aptamer	Average K_D	Chi^2 values
CSIR 2.2	**21.5nM ± 4.3nM**	**0.1**
CSIR 2.11*	**9.0nM ± 1.7nM**	**0.1**
CSIR 2.15	**12.8nM ± 1.6nM**	**1.4**
CSIR2.19	**1.6nM ± 0.5nM**	**0.3**
CSIR 2.21	**10.2nM ± 3.4nM**	**0.3**

The values were obtained using triplicates

* Rotherham et al. (2012)

3.4 Determination of CSIR 2.11-ESAT.CFP-10 dimer interactions by EMSA

EMSA experiments were performed with the aim of assessing the interactions between the aptamers and the ESAT-6.CFP-10 heterodimer, and the CFP-10 monomer, and to potentially characterise the kinetic parameters thereof. For these experiments, the in-house-produced CSIR 2.11 aptamer was used. The aptamer was radioactively labelled at the 5' end with γ-^{32}P, and the labelling was verified by non-denaturing PAGE (Figure 3-11).

The EMSA results did not demonstrate strong binding of the aptamer to the proteins. For the most part, the electrophoresed samples did not show a discernible shift, and in some cases did not form single discrete bands. In certain instances, particularly when lower concentration range of protein was used, two rows of ill-defined bands were observed in the aptamer-protein complex lanes (Figure 3-12, A). The more concentrated (darker) shifted bands were observed towards the bottom end of the gel.

These bands, which represent the free probe (aptamer), ran along the same plane as the band in the aptamer (CSIR 2.11) alone control well. A more diffuse ("smeary") area of shifted material was observed in the top part of the gel, presumably representing aptamer-protein complexes. The CSIR2.11-CFP-10 complexes appear to have migrated less than the CSIR 2.11- ESAT-6.CFP-10 dimer complexes. In the case where a higher protein concentration range was used (Figure 3-12, B), only single solid bands were observed, with no sign of hindered mobility in the aptamer-protein complex lanes compared to the aptamer alone control. Interestingly, the CSIR 2.11-dimer samples migrated further than both the aptamer-CFP-10 and aptamer alone samples. Following several repeats with inconsistent results (Appendix, Figure A-1), the EMSA could not be optimised. The inconsistencies made it difficult to attain reliable qualitative and quantitative data. Thus, the kinetic parameters of the aptamer-ESAT-6.CFP-10 heterodimer could not be determined using this technique.

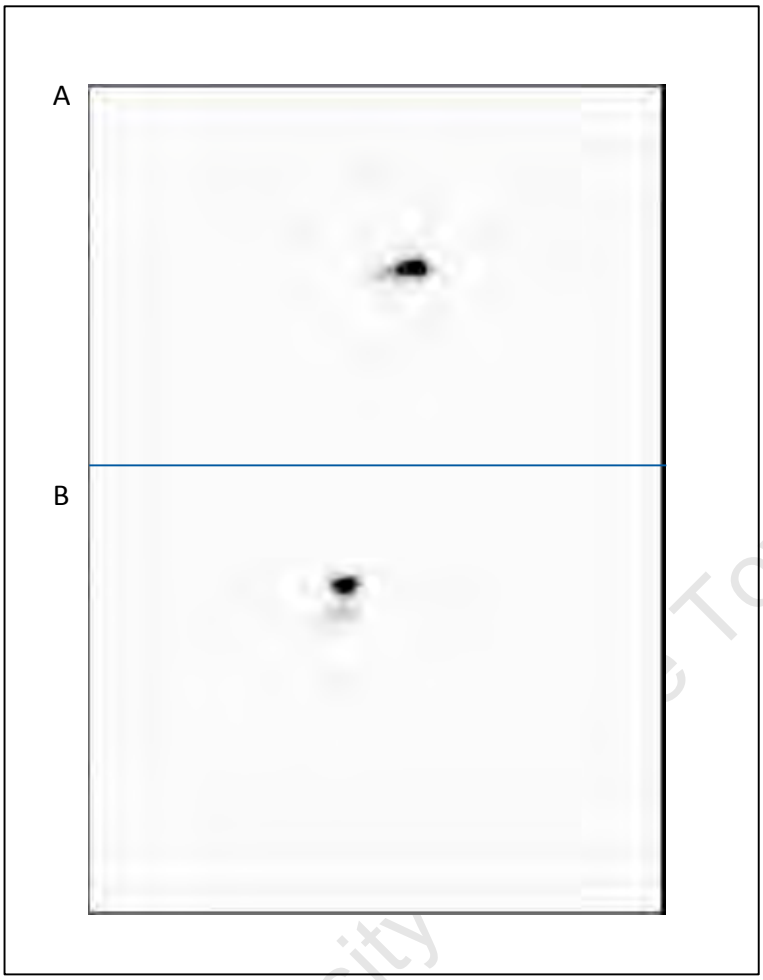


Figure 3-11: Confirmation of radioactive labelling of aptamer by PAGE. Aptamer CSIR 2.11 was labelled with ^{32}P, analysed on denaturing and non-denaturing 8% PAGE gels, and viewed on a phosphorimager. The top panel (A) shows the end-labelled CSIR 2.11 on a denaturing gel, while the bottom panel (B) shows the same aptamer on a non-denaturing gel.

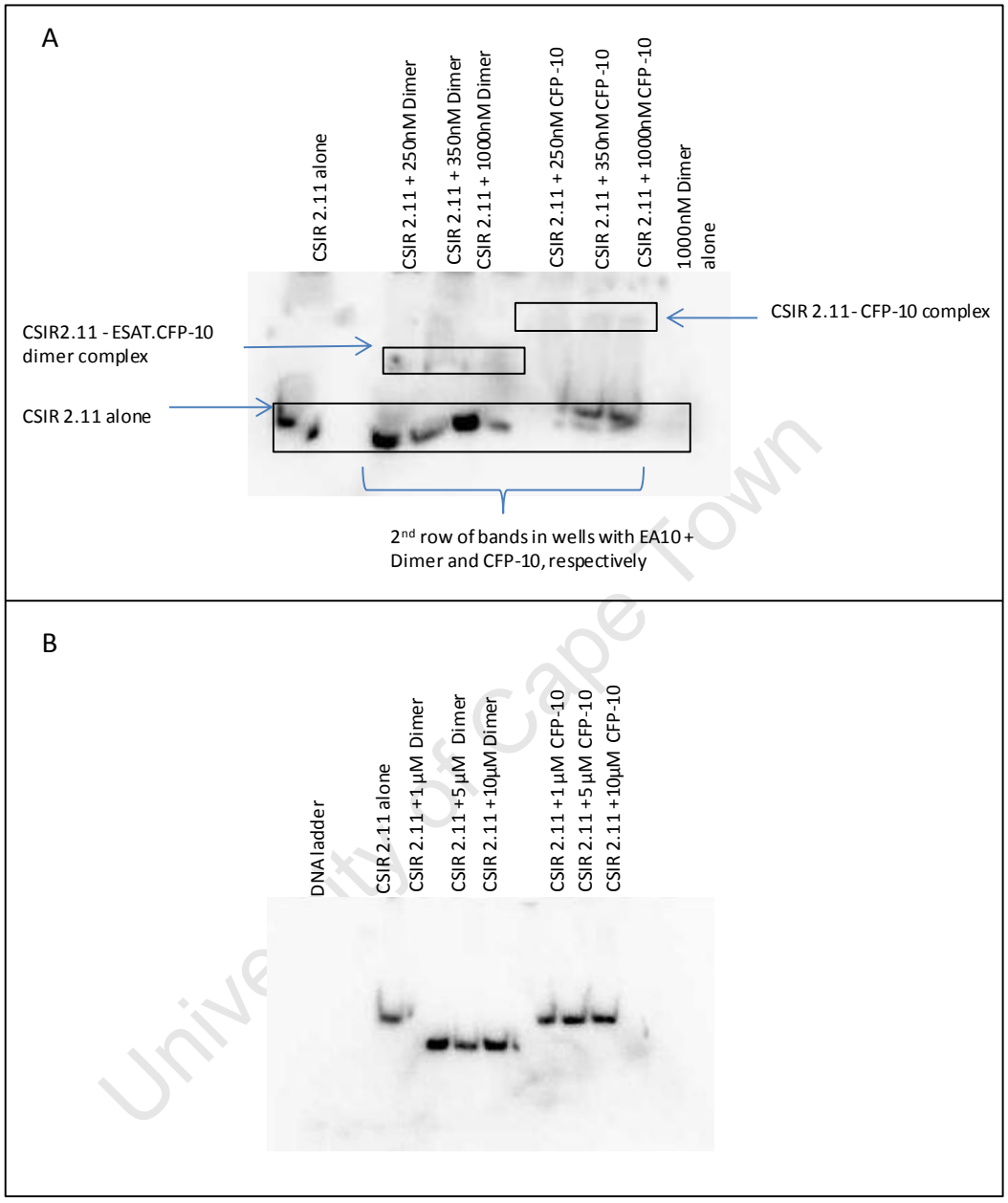


Figure 3-12: Representative gel-shift analysis of the CSIR 2.11 aptamer. Increasing amounts of protein (either CFP-10 or ESAT-6.CFP-10 dimer-indicated above wells) were mixed with ^{32}P-end labelled CSIR 2.11. Samples were run on a native 8% Polyacrylamide gel. The gels were dried and exposed to a phosphor screen and viewed on a phosphorimager. Two sets of protein concentrations were studied. Phosphor images represented in A and B depict the lower and higher protein concentration ranges, respectively.

3.5 Structural characterisation of the full length aptamers

Given that aptamer-target interactions are attributed mainly to structural compatibility, characterisation of the structural features of the aptamers was necessary in order to delineate the structural contributions of these aptamers in target recognition.

3.5.1 Primary structures

The primary structures (sequences) of the six full length aptamers that were chosen for further characterisation are shown in Table 3-2. Sequence alignment using BioEdit Sequence Alignment Editor showed no conserved regions in the random region of the aptamers. There were no common motifs between the aptamers to which their binding functionality could be attributed. The "Plot identities" application of the BioEdit program revealed very little nucleotide sequence similarity between CSIR 2.11 and the five other aptamers (Table 3-3). The dots represent bases are common nucleotides between each of the aptamers and CSIR 2.11. Most of the dots were observed in the constant (primer) regions; which was expected.

Table 3-2: Primary structures (nucleotide sequences) of the Full length aptamers

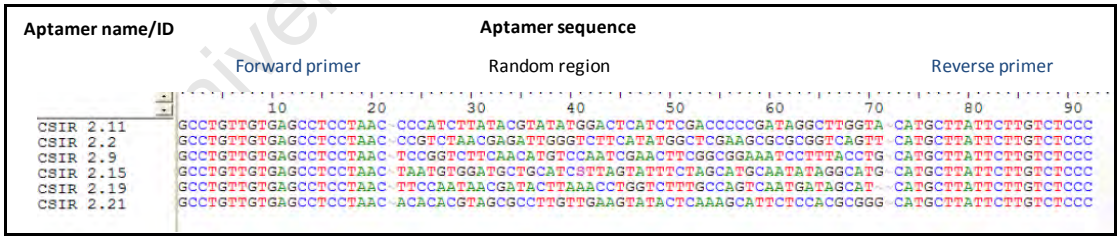
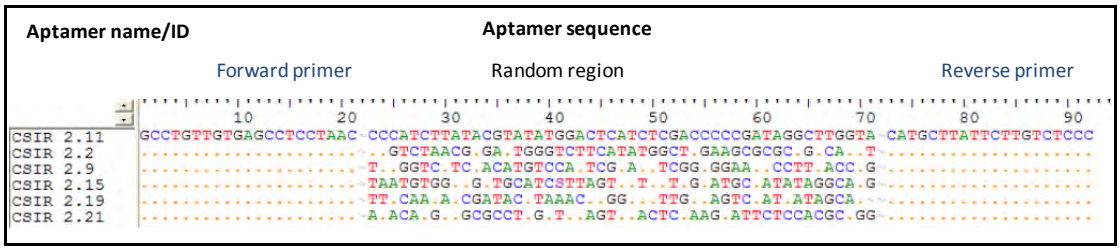


Table 3-3: Plot Identity analysis of the six selected Full length aptamer sequences



3.5.2 Prediction and analysis of secondary structures of full length aptamers using *in silico* methods

Secondary structures of aptamers define which regions of the primary structure contribute to the overall target-binding conformation. Secondary structures of the anti-ESAT.CFP-10 aptamers were predicted with mfold. The program generated for each aptamer, either one or multiple low energy conformations into which the aptamer could stably fold (Figures 3-13 to 3-18). In cases where multiple low energy conformations were predicted, the suboptimal structures showed similar folding patterns in terms of stem-loops, as well as free energy values very close to that of the predicted MFE structure. From the structural simulations, CSIR2.11 and CSIR 2.15 had only one predicted structure under the stated conditions, while CSIR 2.2, CSIR 2.9, CSIR 2.19, and CSIR 2.21 had more than one (Figure 3-13). For CSIR 2.11 and 2.19 in particular, the primer regions, predominantly the 5′ end did not seem to show significant contribution to secondary structures (Figures 3-13 and 3-17).

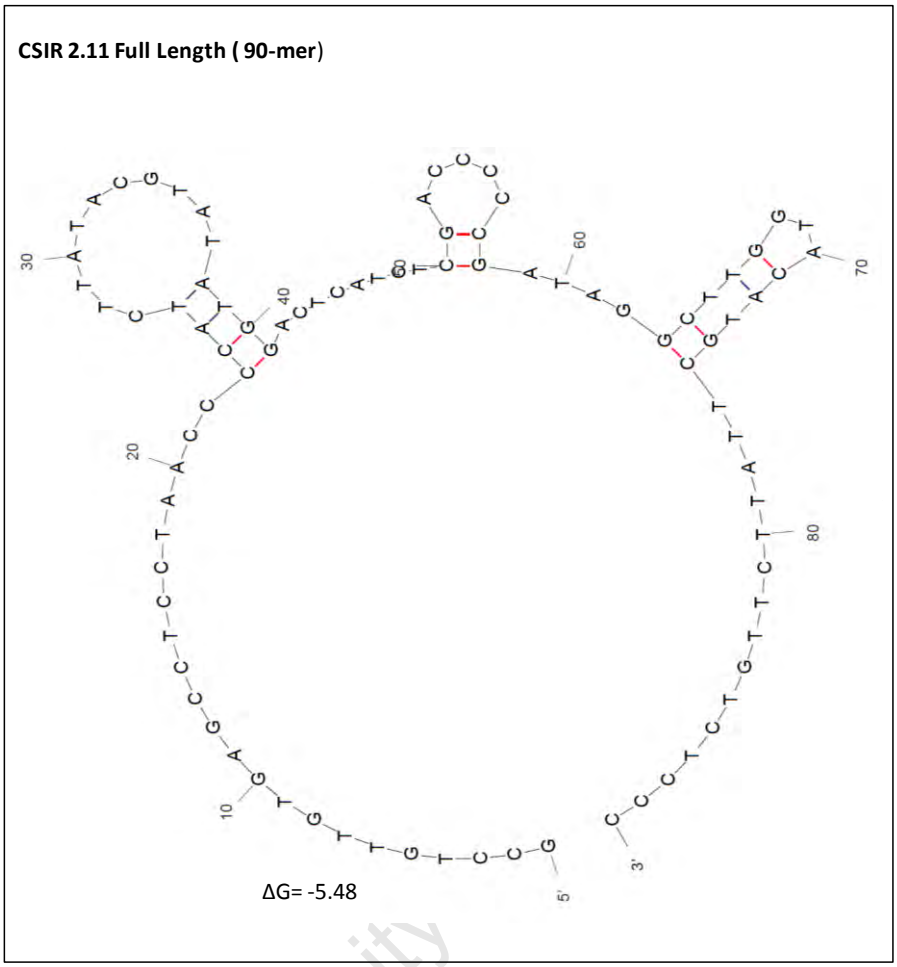


Figure 3-13: Potential secondary structure of aptamer CSIR 2.11 as predicted by mfold. Structure prediction for CSIR 2.11 yielded one minimum free energy (MFE) structure. Positions 1–20 and 68–90 are the primer sequences. The structure shows three stem-loops, with varying nucleotide lengths, which stabilize the structure of the aptamer. The ΔG represents the free energy (in kcal/mol) of the predicted conformation.

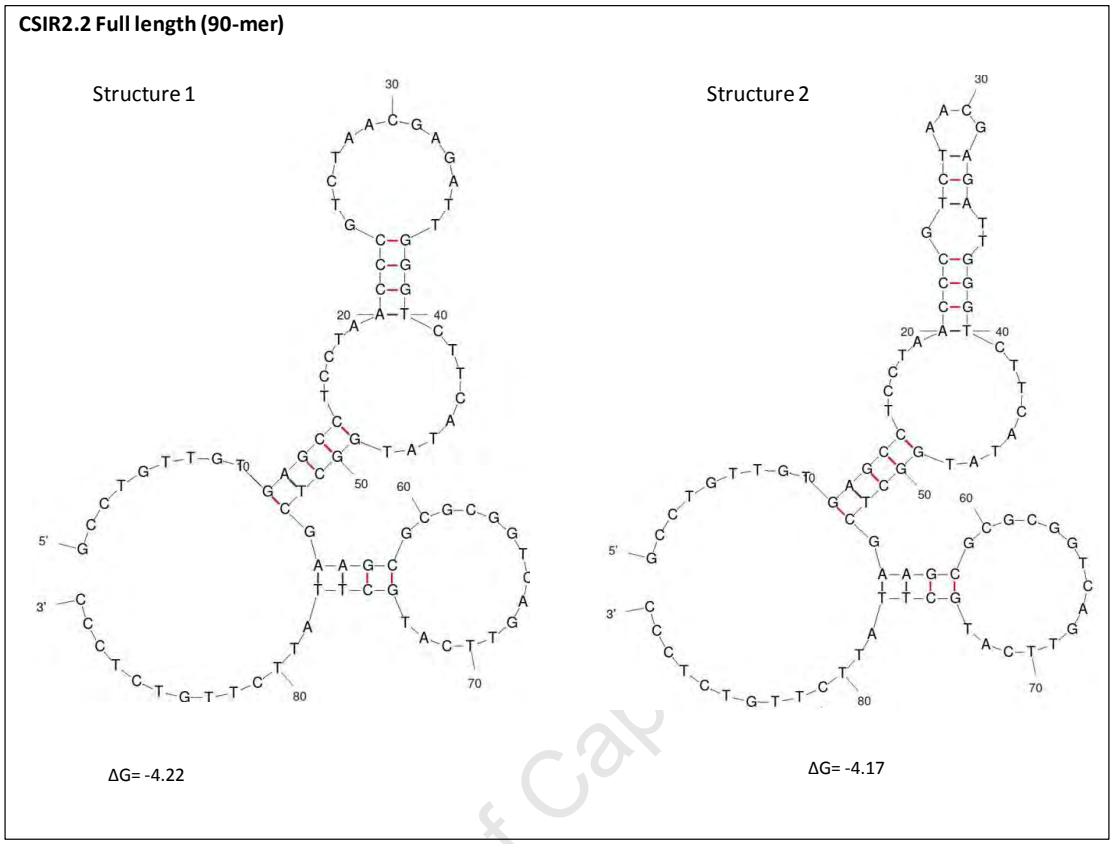


Figure 3-14: Potential Secondary Structures of aptamer CSIR 2.2 as predicted by mfold. Structure prediction for CSIR 2.2 yielded two structures within close free energy range. Positions 1–20 and 68–90 are the primer sequences. The structure shows stem-loops, with varying nucleotide lengths, which stabilize the structure of the aptamer. The ΔG represents the free energy (in kcal/mol) of the predicted conformation.

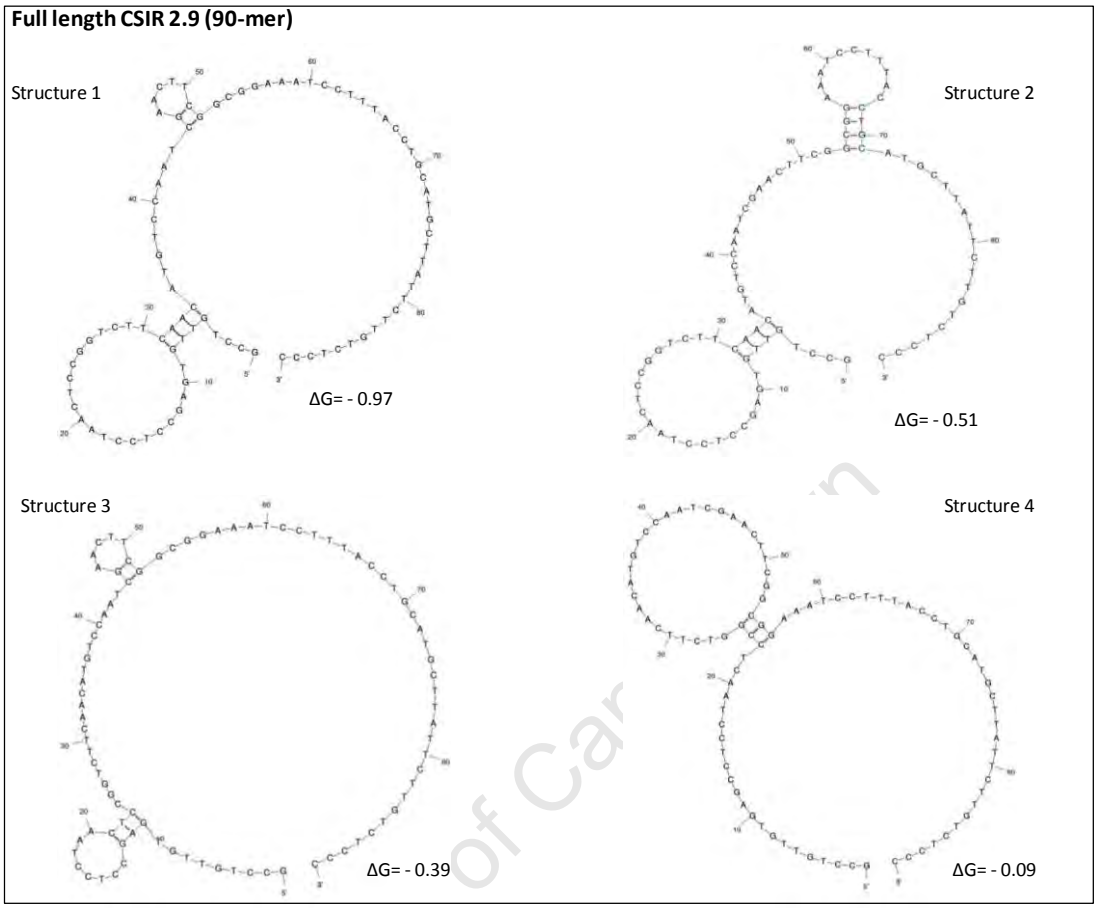


Figure 3-15: Potential Secondary Structures of aptamer CSIR 2.9 as predicted by mfold. Structure prediction for CSIR 2.9 yielded four minimum free energy structures. Positions 1–20 and 68–90 are the primer sequences. The structure shows stem-loops, with varying nucleotide lengths, which stabilize the structure of the aptamer. The ΔG represents the free energy (in kcal/mol) of the predicted conformation.

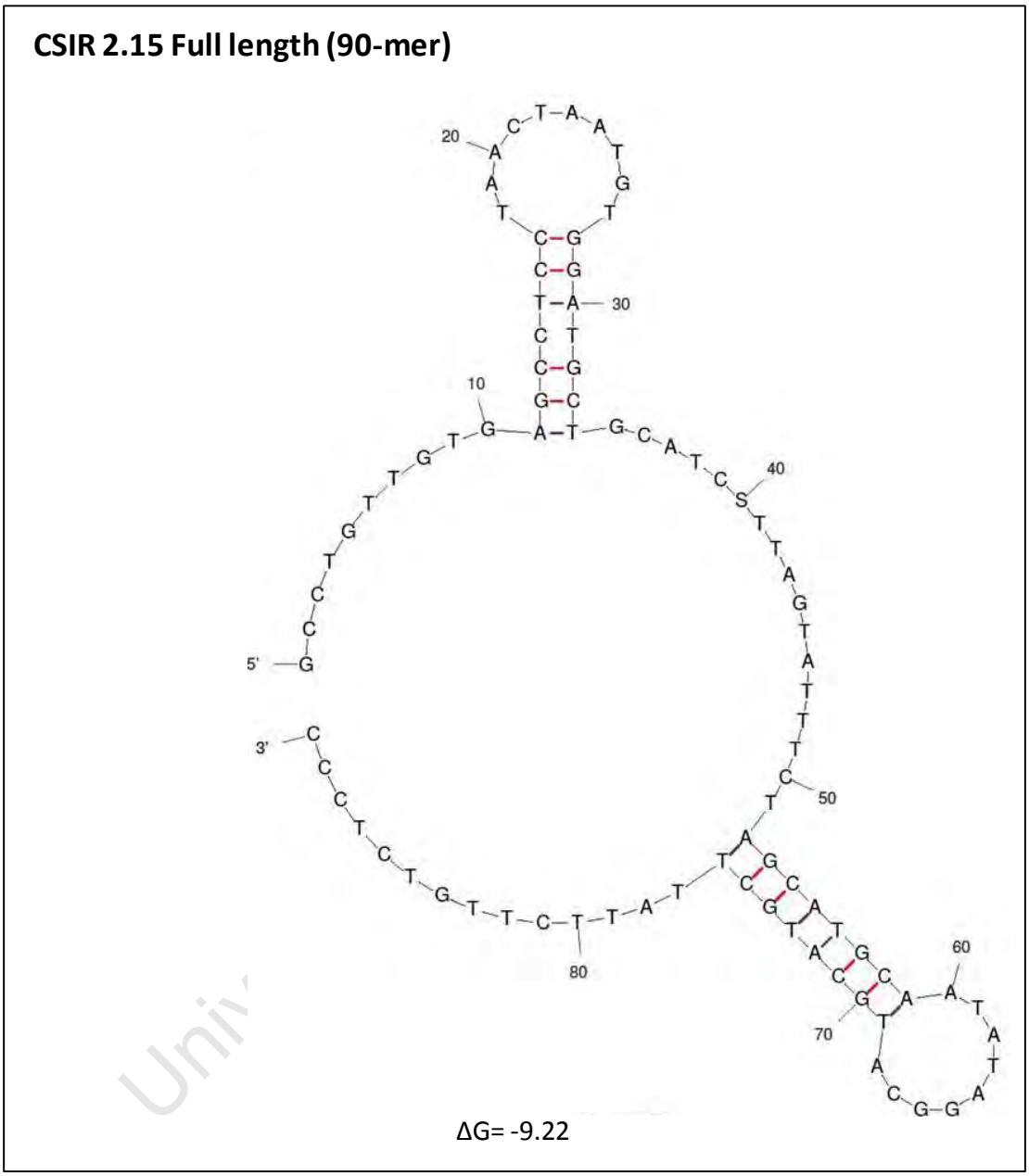


Figure 3-16: Potential Secondary Structure of aptamer CSIR 2.15 as predicted by mfold. Structure prediction for CSIR 2.15 yielded one minimum free energy structure. Positions 1–20 and 68–90 are the primer sequences. The structure shows stem-loops, with varying nucleotide lengths, which stabilize the structure of the aptamer. The ΔG represents the free energy (in kcal/mol) of the predicted conformation.

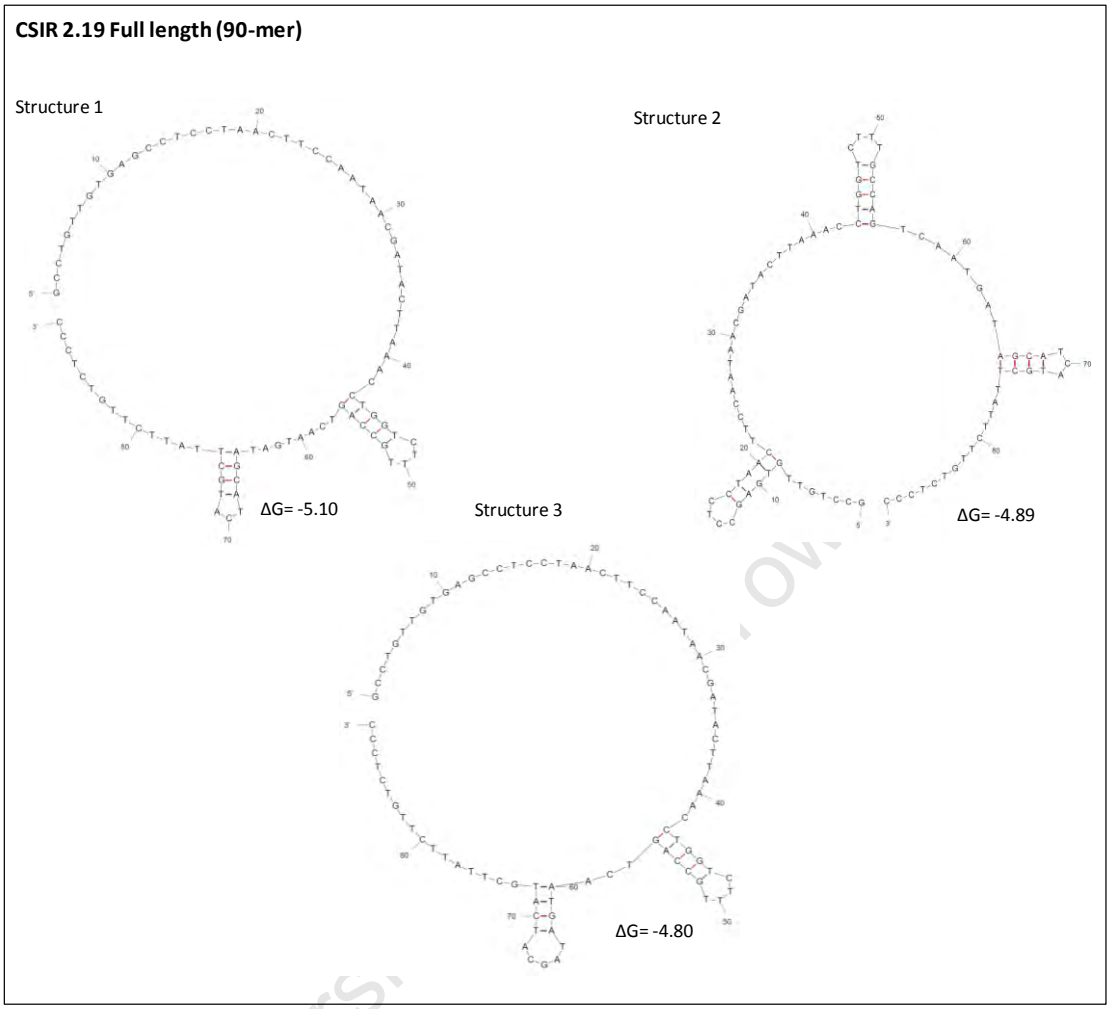


Figure 3-17: Potential Secondary Structures of aptamer CSIR 2.19 as predicted by mfold. Structure prediction for CSIR 2.19 yielded three minimum free energy structures. Positions 1–20 and 68–90 are the primer sequences. The structure shows stem-loops, with varying nucleotide lengths, which stabilize the structure of the aptamer. The ΔG represents the free energy (in kcal/mol) of the predicted conformation.

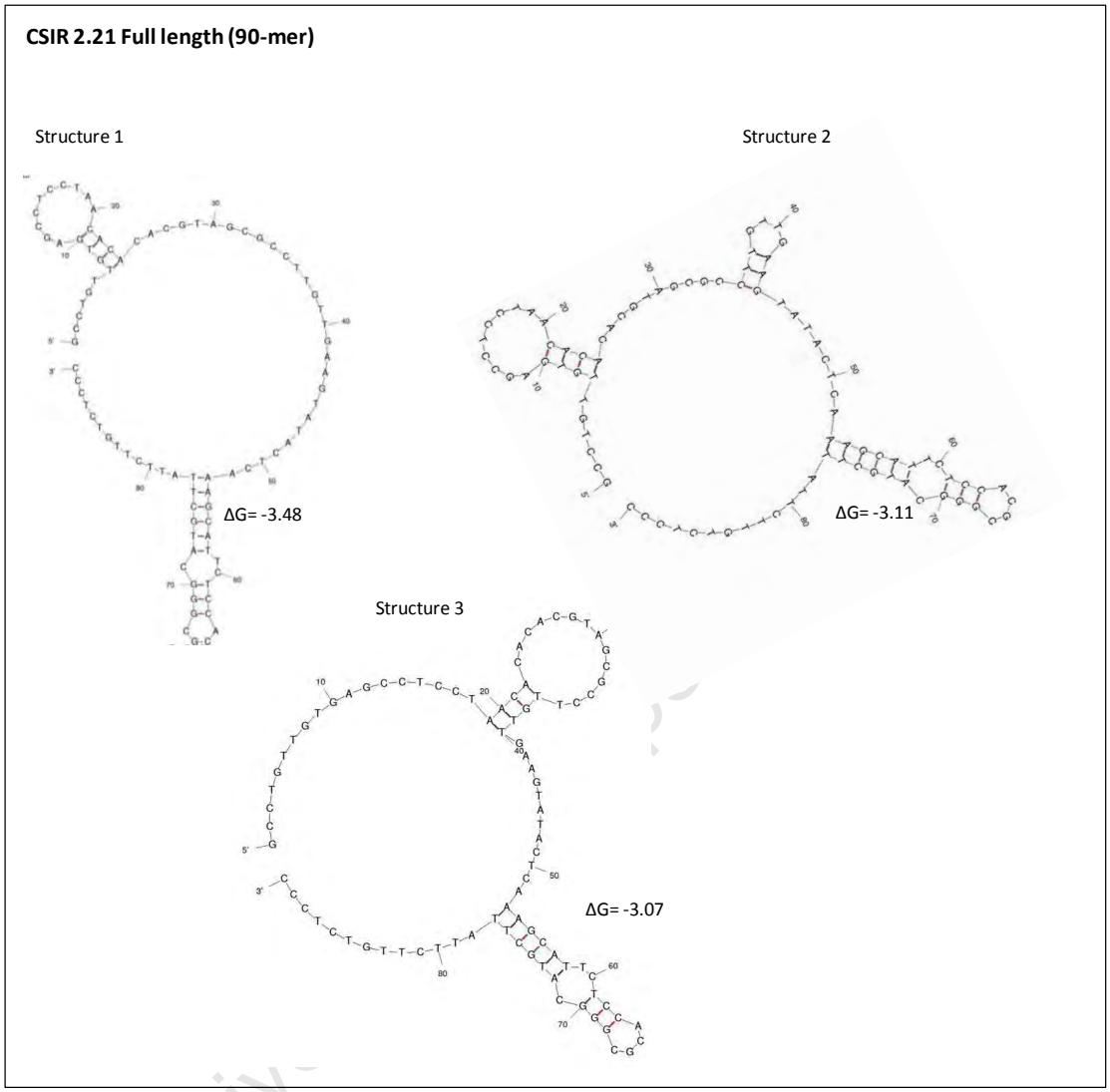


Figure 3-18: Potential Secondary Structures of aptamer CSIR 2.21 as predicted by mfold. Structure prediction for CSIR 2.21 yielded three minimum free energy structures. Positions1–20 and 68–90 are the primer sequences. The structure shows stem-loops, with varying nucleotide lengths, which stabilize the structure of the aptamer. The ΔG represents the free energy (in kcal/mol) of the predicted conformation.

3.6 Rational truncation of full length aptamers based on predicted secondary structures

Based on the secondary structure analyses, the nucleotides comprising the stem-loop combinations of the aptamer structures were hypothesized to be the key elements involved in target-binding. Therefore, rational truncation of CSIR 2.11 and 2.19 was carried out by deleting parts of the full length sequences, primarily the nucleotides in the constant regions. The aptamers were truncated as much as possible without affecting their final secondary structures. This process yielded shorter versions of the aptamers which maintained the overall structure of the predicted target-binding regions of the parent full length aptamers (Figures 3-19 and 3-20).

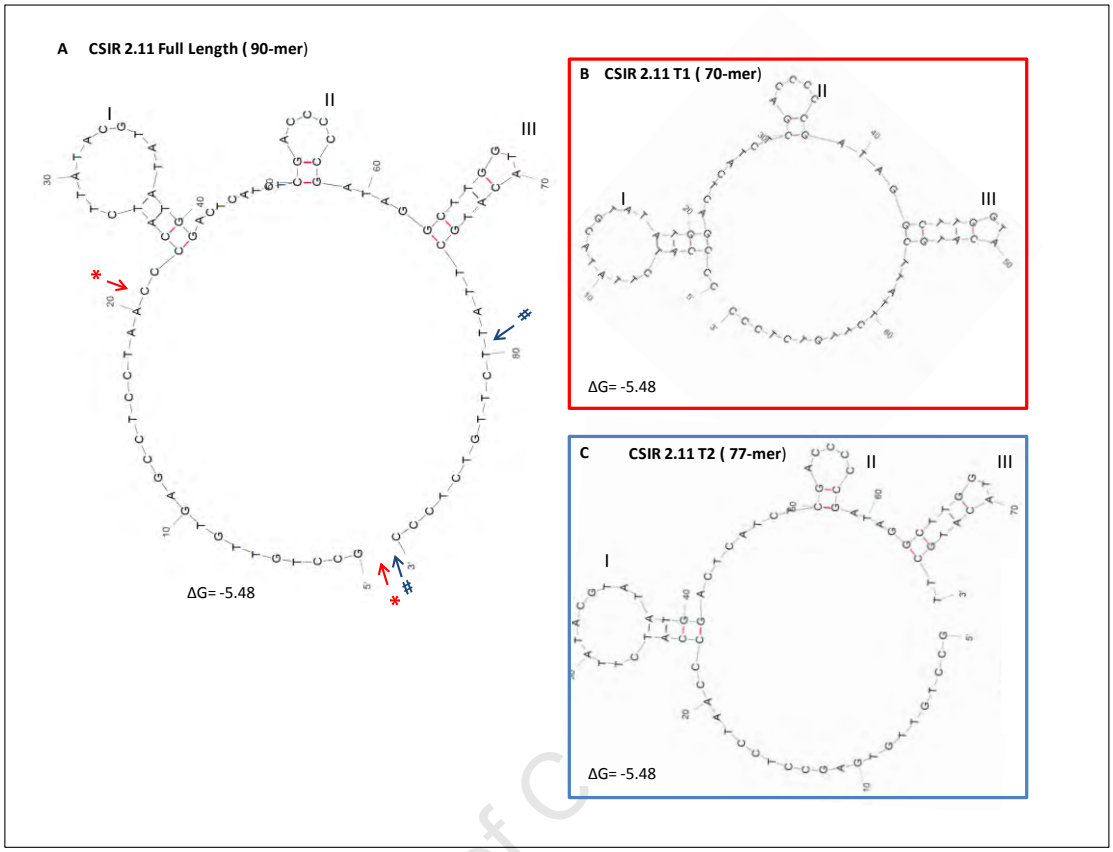


Figure 3-19: Truncation of aptamer CSIR 2.11 guided by its secondary structure. A. The predicted secondary structure for the original sequence of aptamer CSIR 2.11 consists of three stem-loops (I to III). All three stems consist of standard Watson-Crick base pairs. The numbers of number of base pairs in each stem are four, two, and five respectively. Truncation 1 (T1), in B (70-mer) resulted from removing 20 nucleotide bases in the direction the 5′ to 3′ (arrows, *), while the second truncation, T2, shown in C (77-mer) resulted from cutting out 13 nucleotide bases in the direction 3′ to 5′ (arrows, #). All the stem-loops, along with their assigned free energy (ΔG, in kcal/mol), were retained in the mfold-predicted structures of both truncated versions of the aptamer.

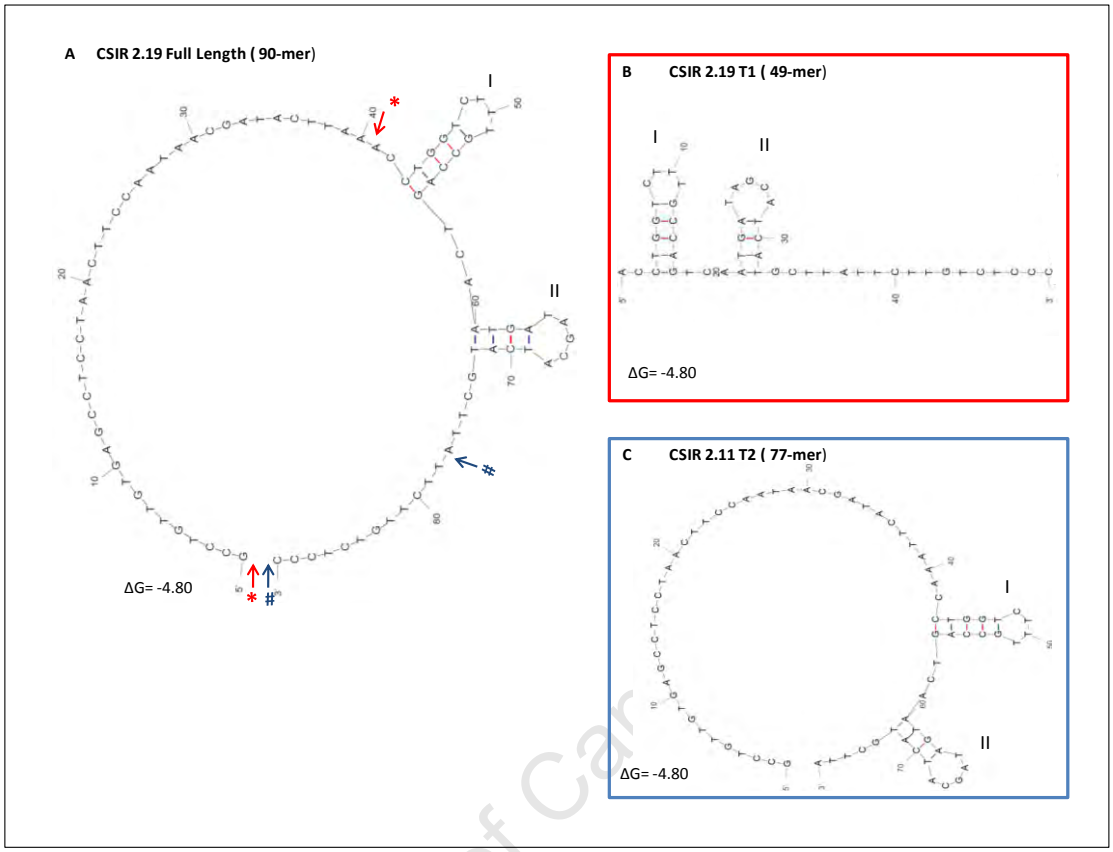


Figure 3-20: Truncation of aptamer CSIR 2.19 by secondary structure-guided methods. A. The predicted secondary structure for the original sequence of aptamer CSIR 2.19 consists of two stem-loops (I to II). Both stems consist of standard Watson-Crick base pairs. The numbers of number of base pairs in each stem are five, and four, respectively. Truncation 1 (T1), in B (49-mer) resulted from removing 41 nucleotide bases in the direction the 5′ to 3′ (arrows, *), while the second truncation, T2, shown in C (77-mer) resulted from cutting out 13 nucleotide bases in the direction 3′ to 5′ (arrows, #). Both stem-loops, along with their assigned free energy values (ΔG, in kcal/mol), were retained in the mfold-predicted secondary structures of both truncated versions of the aptamer.

3.7 Conformational energy landscape analysis of full length aptamers and truncated derivatives

3.7.1 *In silico* analysis of folding behaviour of CSIR 2.11 and 2.19: energetics and folding kinetics

Conformational energy landscapes of CSIR 2.11 and CSIR 2.19 were simulated using the *Barriers program* of the Vienna RNA Web server. Barrier trees and folding kinetics simulations of aptamers CSIR 2.11 and CSIR 2.19 and their truncated derivatives are illustrated (Figures 3-21 to 3-26). While the program computes folding landscapes over millions of possible conformations for one sequence, for simplicity, only the 50 lowest local minima are shown in the barrier tree for each aptamer. The energy barriers between these minima are presented above each leaf. Folding kinetics simulations were computed by *treekin* using the 50 local minima (represented on each tree) as macro-states for the coarse-grained dynamics simulations. The *treekin* output is represented as folding kinetics graphs (Figures 3-21 to 3-26—on the right hand side of each barrier tree). These graphs show the fraction of folding pathways for each aptamer sequence en route to the MFE structure, plotted on a logarithmic time scale. The time scale is represented in arbitrary units, referred to as "time steps". The graphs portray the population density of the basins of attraction of the local minima as a function of time. This gives an indication of which structures are likely to compete with the MFE and possibly hinder the interaction of the target with the MFE, should the MFE not be reached within a biologically reasonable time frame. The folding process was started in the open chain state equivalent (indicated in black on the graphs for each of the aptamers analysed), and run until thermodynamic equilibrium distribution was reached within the user-specified time-frame.

With the folding process simulations presented in herein, the focus was not on the time it takes to reach a thermodynamically stable equilibrium distribution, but rather the refolding time from a selected start structure (representing the denatured or open chain structure) to the MFE structure (the global minimum, denoted by the number 1 in each tree, and represented in red on each folding kinetics graphs).

3.7.2 Conformational energy landscapes of full length and truncated derivatives of CSIR 2.11

For the CSIR 2.11 90-mer aptamer**,** the tree shows three distinct subtrees, which represent the three most dominant folding pathways (Figure 3-21). The basins of attraction (seen in the three subtrees) converge into structures 1, 4 and 33, in respective order of increasing free energy. Structures 1, 4 and 33 are the lowest hanging minima in each basin of attraction or subtree. The subtree that is the most occupied (i.e. with the most leaves), and the more likely kinetically preferred, is the one containing the local minimum 4 and not the MFE (1) as would have been expected. However, once the MFE is reached, it is not likely for aptamer CSIR 2.11 to switch between conformations in the two subtrees. This is due to the fact that in terms of barrier heights, the energy landscape portrays a downhill walk from 4 to 1, while the opposite direction is something of an uphill climb.

Also shown in Figure 3-21 (B) is a graphical representation of the folding kinetics of the aptamer as determined by the treekin simulations. As indicated, the graph shows the fraction of folding pathways leading to the MFE, and the population density of the basins of attraction as a function of time. For the 90-mer CSIR 2.11 tree, the denatured or open-chain conformation was represented by local minimum 50 (the structure in the ensemble with a ΔG value closest to zero), which is directly connected to local minima 44 and 33 via energy barriers of 0.4 and 0.5kcal/mol, respectively. The simulation showed an initial rapid decline in the population of local minimum 50 (the "open chain"); from 100 percent to approximately zero before 1000 time-steps. As expected (based on the barrier tree), local minima 33 and 44, both of which belong to the same basin of attraction as local minima 50, were dominant in the beginning of the refolding process. The two were later followed by local minima 4 and 8, after which the simulation ended (within the set time frame of ~ 1e+10) with the MFE coming up to less than 20% population density. Other local minima that may have existed in the population are not shown in the graph, as they were populated at rates of less than 10%, which was the chosen minimum in the user-specified treekin settings.

The 77-mer CSIR 2.11 showed a similar folding pattern to that of the 90-mer, both on the tree and associated kinetics, with the MFE being reached only after time steps 1e+03 (Figure 3- 22). The basins of attraction, as seen in the three subtrees, converge into structures 1, 5 and 22, in respective order of increasing free energy. The most populated subtree contains local minima 5, 6, and 7, which may act as kinetic traps that could hinder the folding path to the MFE. However, the thermodynamic cost of switching between conformations in this subtree and the one containing the MFE is high (as indicated by the barrier height). Furthermore, the MFE population of the 77-mer is slightly higher than that of the 90-mer and the 70-mer.

The 70-mer showed a different folding pattern to 90-mer and 77-mer, with the subtree that is most populated containing the MFE and the population of the MFE peaking earlier, at time step 1e+02 (Figure 4-23). The basins of attraction, as seen in the three subtrees, converge into structures 1, 5, 36 and 46, in respective order of increasing free energy. These structures are the lowest hanging minima. The subtree that is the most occupied and the more likely kinetically preferred contains the MFE. The significant local minima, such as 5, 8, 36 and 46 are not likely to act as kinetic traps. Local minimum 2 appears to be a strong competitor of the MFE, as the two structures show up and reach almost equal maximum density around the same time. This suggests a likely conformation switch between the two structures. However, the MFE would probably be more thermodynamically preferred as it has a lower ΔG.

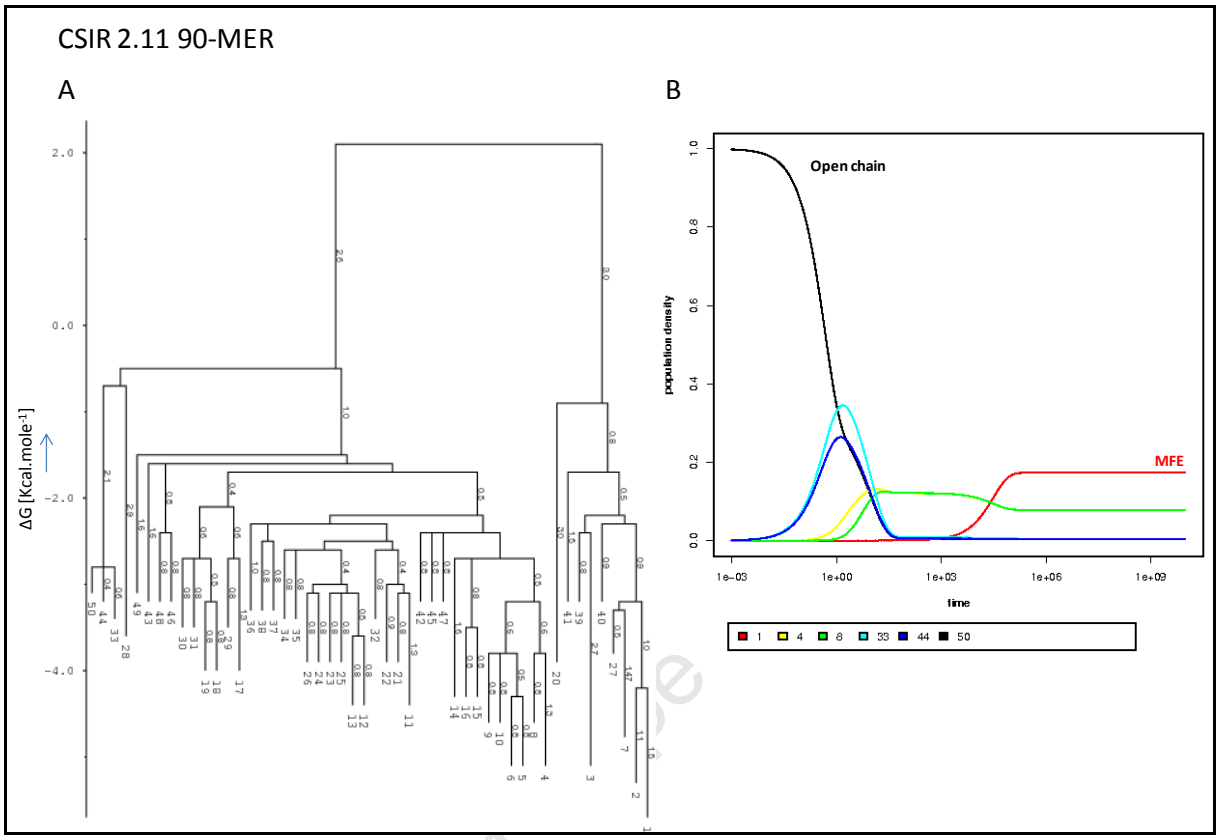


Figure 3-21: Barrier tree and folding kinetics of the 90-mer CSIR 2.11. **A**. The barrier tree the lower energy portion of the conformational landscape. The leaves (1 to 50) of the tree represent the 50 lowest local minima numbered by their energy, and the numbers labelling the branches are the barrier heights denoting the free energy differences between the local minima) in kilocalories/mol. The saddle points (nodes) are the energetically highest points on a folding path between any two local minima. The basins of attraction (seen in the three subtrees) converge into structures 1, 4 and 33, in respective order of increasing free energy. **B**. Folding kinetics of CSIR 2.11 from the open chain structure (computed by the treekin program of the Vienna RNA Barriers). The plot shows the population density of meta-stable states along a simulated folding path beginning from the open chain (or the structure with the highest free energy value, in this case structure 50, black line), and eventually reaching the MFE structure. The population density of the initial state (structure 50) quickly declines, reaching 50% at time step 1e+ 00, and later 0% by time 1e+03. Local minima 44 and 33 appear more dominant at the beginning of the refolding process, reaching 25 and 35 % density at time 1e+00, respectively. The global minimum (MFE—structure 1) reaches its highest density later at time step 1e+05, and stabilizes at a density of about 18%, with local minima 4 and 8 (belonging to the same basin) just below it

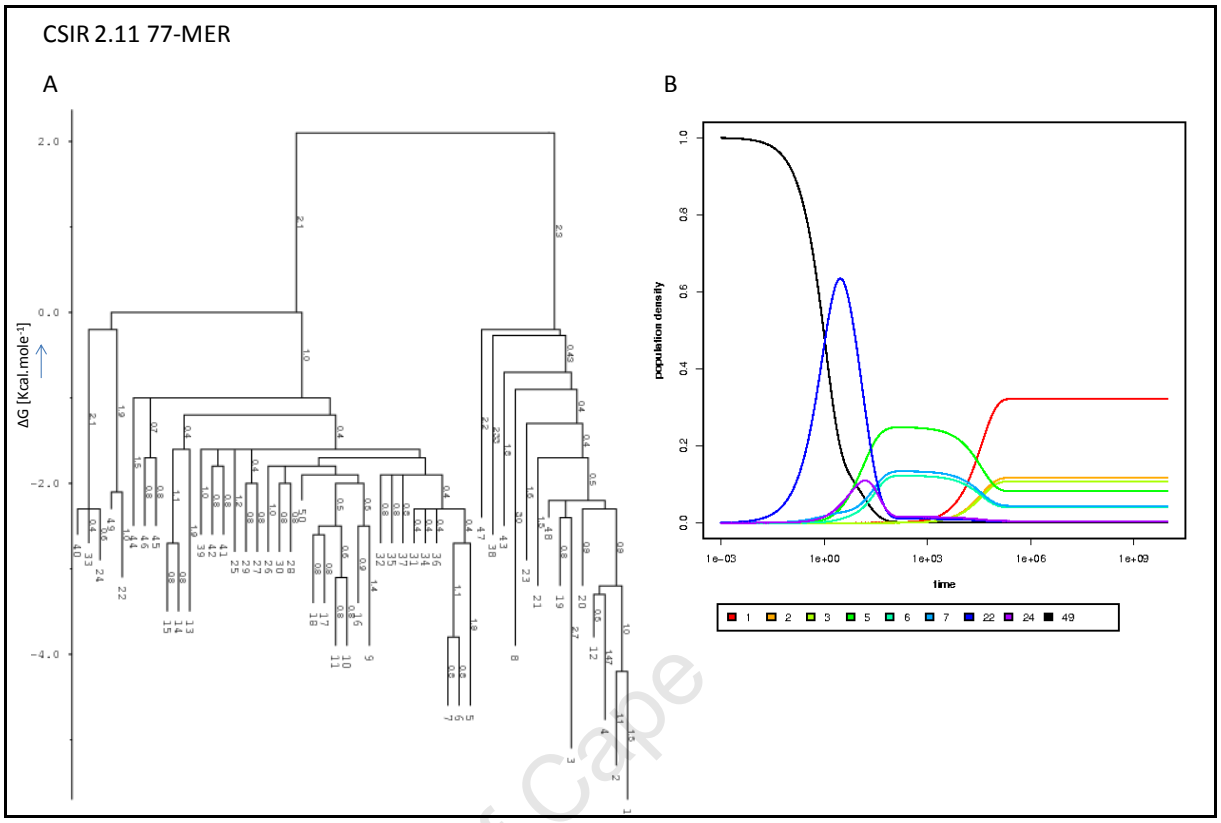


Figure 3-22: Barrier tree and folding kinetics of the 77-mer CSIR 2.11. The barrier tree is shown in **A.** The numbers 1 to 50 represent the 50 lowest local minima numbered by their energy, and the numbers labelling the branches are the barrier heights denoting the free energy differences between the local minima) in kilocalories/mole. The saddle points (nodes) are the energetically highest points on a folding path between any two local minima. There are two distinct folding pathways, with the basins of attraction (seen in the two subtrees) converging into structures 1, and then 5, 22, and 24 in respective order of increasing free energy (5, 22, and 24 are found in the same subtree). **B.** folding kinetics of CSIR 2.11 77-mer from the open chain structure (computed by the treekin program of the Vienna RNA Barriers). The plot shows the population density of meta-stable states along a simulated folding path beginning from the open chain (or the structure with the highest free energy value, in this case structure 49, black line) and eventually reaching the MFE conformation. The population density of the initial state (structure 49) quickly declines, reaching 50% at time step 1e+ 00, and later 0% before 1e+03. Local minima 22 appears more dominant at the beginning of the refolding process, reaching 60 % density by time 1e+00, quickly followed by 24, and then local minima 5, 6 and 7 which are found in the same basin of attraction. Local minima 2 and 3 start to show up in the population around the same time as the global minimum (MFE—structure 1), which reaches its highest density and dominance later just after time step 1e+04, and stabilizes at a density of over 30%.

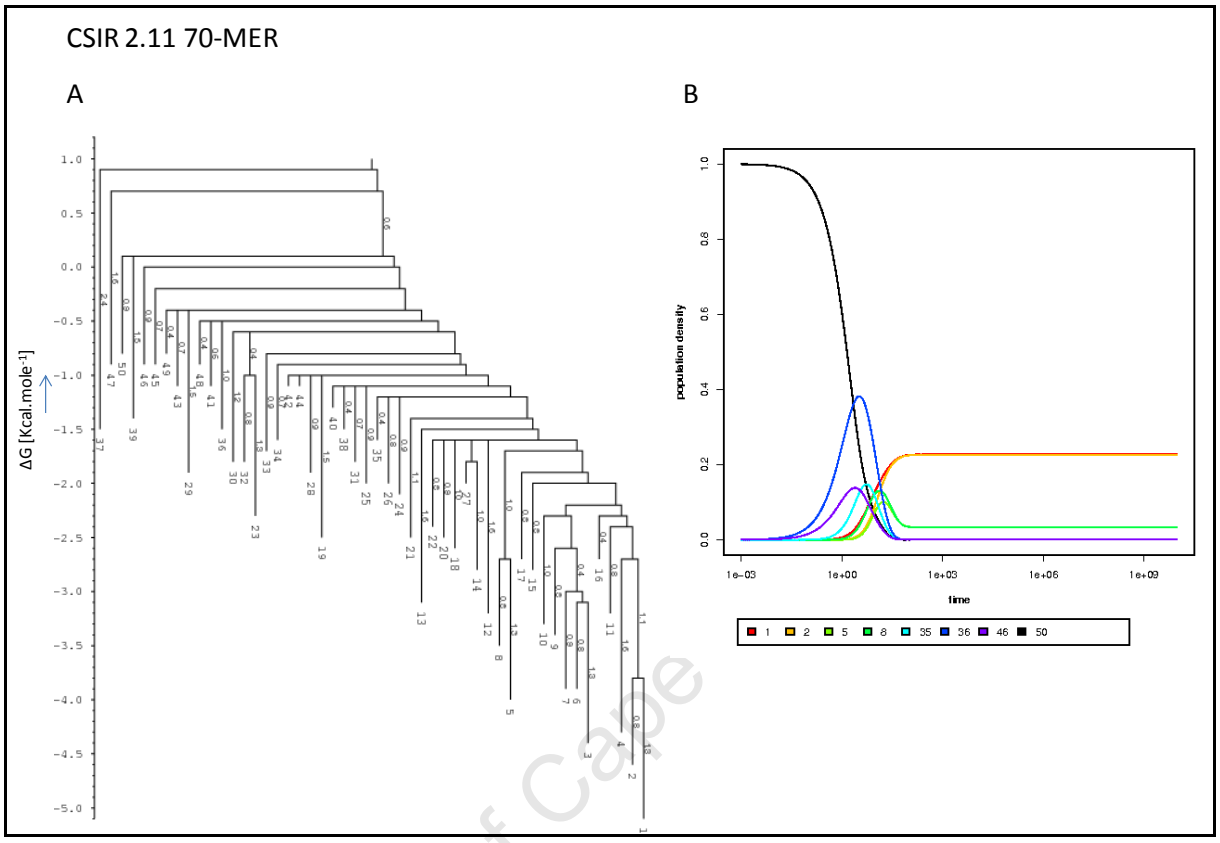


Figure 3-23: Barrier tree and folding kinetics of the 70-mer CSIR 2.11. The barrier tree is shown in **A**. The numbers 1 to 50 represent the 50 lowest local minima numbered by their energy, and the numbers labelling the branches are the barrier heights denoting the free energy differences between the local minima) in kilocalories/mole. The saddle points (nodes) are the energetically highest points on a folding path between any two local minima. There are two distinct folding pathways, represented by two basins of attraction: the one converging into structure 1 and that containing structures 36 and 46. **B**. Folding kinetics of CSIR 2.11 70-mer from the open chain structure (computed by the treekin program of the Vienna RNA Barriers). The plot shows the population density of meta-stable states along a simulated folding path beginning from the open chain (or the structure with the highest free energy value, in this case structure 50, black line) and eventually reaching the MFE conformation. The population density of the initial state (structure 50—black) quickly declines, reaching 50% at time step 1e+ 00, and later 0% before 1e+03. Local minima 36 appears more dominant at the beginning of the refolding process, reaching almost 40 % density by time 1e+00, quickly followed by 46, 35, and 5 and 8 which are found in the same basin of attraction. The global minimum (MFE—structure 1) makes up part of the population at time step 1e+00 already, competing with structures 5, 8, and 2, and peaks just after time step 1e+02, and stabilizes at a density of over 20%, with local minima 2 (belonging to the same basin) just almost at the same density.

3.7.3 Conformational energy landscapes of full length and truncated derivatives of CSIR 2.19

For the 90-mer CSIR 2.19, the major basins of attraction converge into structures 1, 2, and 16 in respective order of increasing free energy (Figure 3-24). The subtree that is the most occupied and the more likely kinetically preferred is the one containing the MFE (1). This is a more favourable picture of the folding energy surface. Local minimum 16 starts out as the dominant structure, with the highest occupancy of over 60%. Local minimum 2 is a strong competitor of the MFE, which again, implies a likely conformation switch between the two structures, more so because the two have equal ΔG values. However the percentage occupancy of the MFE peaks earlier than that of structure 2. This suggests that the MFE will have some time to interact with the target before structure 2 has a chance compete.

The 49-mer CSIR 2.19 aptamer showed similar folding kinetics to that of the 90-mer, but with a higher MFE population of almost double that of the 90-mer (Figure 3-25). The basins of attraction converge more prominently into structures 1, 2, 3, 12 and 31, in respective order of increasing free energy. Structure 1 and 2 are strong thermodynamic competitors as depicted on the tree and the folding kinetics graph. However, the percentage occupancy of structure 2 only reaches about half that of 1 after time step 1e+ 03. This suggests that structure 1 (MFE) is the more kinetically preferred structure. The two structures have maximum percentage occupancy in the same range—an expected outcome since they have almost equal free energy at thermodynamic equilibrium, even though they belong to different basins of attraction.

The 77-mer CSIR 2.19-mer showed favourable kinetics, with the most occupied subtree containing the MFE (Figure 3-26). The MFE population of the 77-mer aptamer is higher than that of the 90-mer and 49-mer and is reached slightly earlier, before time step 1e+ 02. The MFE and local minimum 2 are found in different basins in this case. This favours the MFE in that once formed, the aptamer is not likely to switch to local minima 2 conformation because of the high barrier it will have to traverse to get to 2.

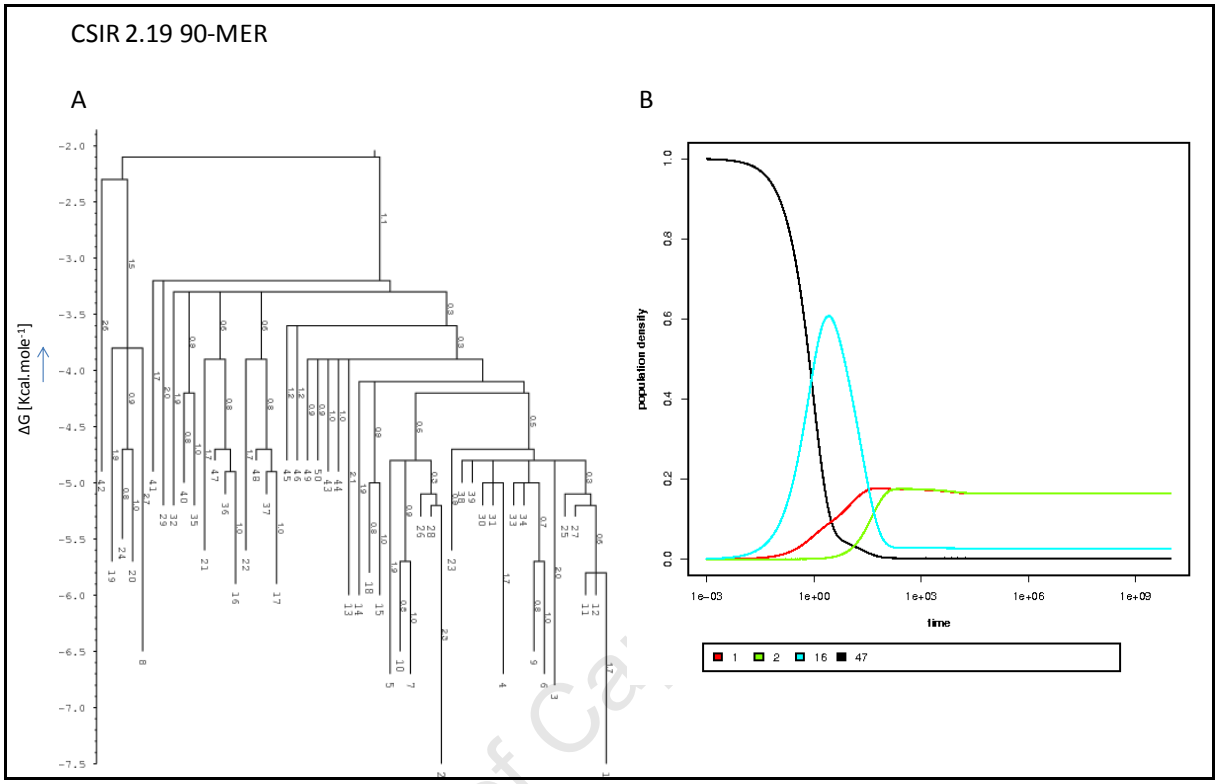


Figure 3-24: Barrier tree and folding kinetics of the 90-mer CSIR 2.19. The barrier tree is shown in **A.** The numbers 1 to 50 represent the 50 lowest local minima numbered by their energy, and the numbers labelling the branches are the barrier heights denoting the free energy differences between the local minima) in kilocalories/mole. The saddle points (nodes) are the energetically highest points on a folding path between any two local minima. There are three distinct folding pathways, with the basins of attraction (seen in the three subtrees) converging into structures 1, 2, and 16 in respective order of increasing free energy. **B.** folding kinetics of CSIR 2.19 90-mer from the open chain structure (computed by the treekin program of the Vienna RNA Barriers). The plot shows the population density of meta-stable states along a simulated folding path beginning from the open chain (or the structure with the highest free energy value, in this case structure 47, black line) and eventually reaching the MFE conformation. The population density of the initial state (structure 47—black) quickly declines, reaching 50% at time step 1e+00, and later 0% before 1e+03. Local minima 16 appears more dominant at the beginning of the refolding process, reaching 60 % density by time 1e+00, quickly followed by the global minimum (MFE—structure 1) and local minimum 2, which is a strong competitor of structure 1 . Structure 1 peaks earlier, structure 2 catches up shortly after 1, and the two share similar density—an expected outcome since they have almost equal free energy at thermodynamic equilibrium, even though they belong to different basins of attraction.

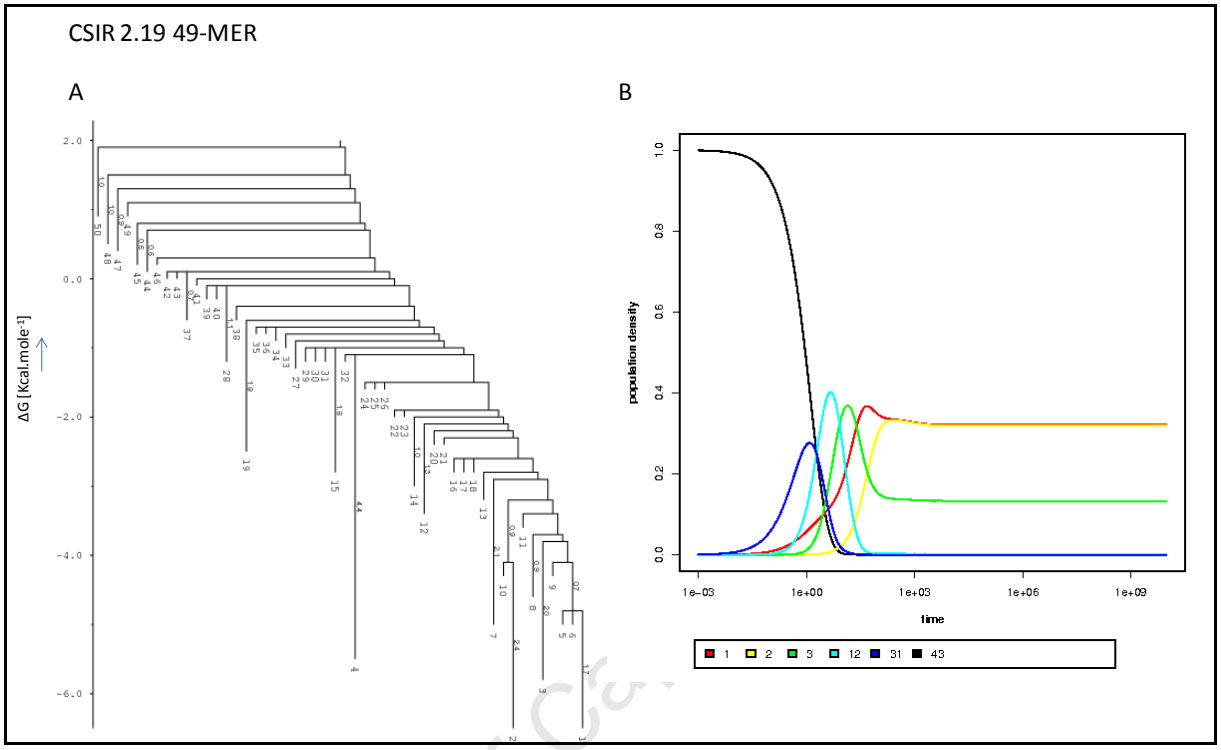


Figure 3-25: Barrier tree and folding kinetics of the 49-mer CSIR 2.19. The barrier tree is shown in **A**. The numbers 1 to 50 represent the 50 lowest local minima numbered by their energy, and the numbers labelling the branches are the barrier heights denoting the free energy differences between the local minima) in kilocalories/mole. The saddle points (nodes) are the energetically highest points on a folding path between any two local minima. There are three distinct folding pathways, with the basins of attraction (seen in the three subtrees) converging into structures 1 and 3 (same basin), then 2, 12, and 31, in respective order of increasing free energy. **B**. folding kinetics of the 49-mer CSIR 2.19 from the open chain structure (computed by the treekin program of the Vienna RNA Barriers). The plot shows the population density of meta-stable states along a simulated folding path beginning from the open chain (or the structure with the highest free energy value, in this case structure 43, black line), and eventually reaching the MFE conformation. The population density of the initial state (structure 43, at ΔG~0) quickly declines, reaches 50% at time step 1e+ 00, and later drops to 0% before 1e+03. Local minima 31 appears more dominant at the beginning of the refolding process, reaching over 40 % density by time 1e+00, quickly followed by 12, 3, the global minimum (MFE—structure 1) and local minimum 2, which is also a strong competitor of structure 1 . Structure 1 peaks earlier than structure 2, and the two share similar density—an expected outcome since they have almost equal free energy at thermodynamic equilibrium, even though they belong to different basins of attraction.

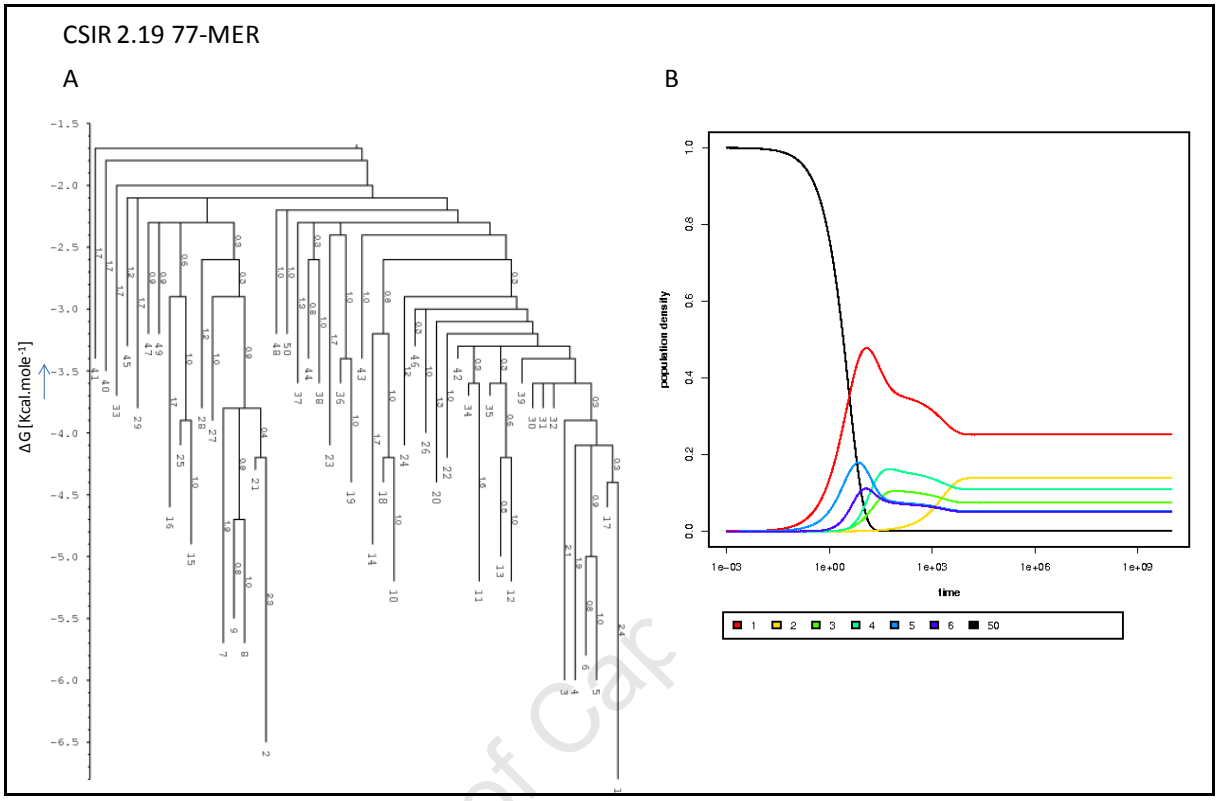


Figure 3-26: Barrier tree and folding kinetics of the 77-mer CSIR 2.19. The barrier tree is shown in **A.** The numbers 1 to 50 represent the 50 lowest local minima numbered by their energy, and the numbers labelling the branches are the barrier heights denoting the free energy differences between the local minima) in kilocalories/mole. The saddle points (nodes) are the energetically highest points on a folding path between any two local minima. There are two distinct folding pathways, with the basins of attraction (seen in the two subtrees) converging into structures 1 (with 3, 4, 5 and 6 in the same basin) and 2 (different basin), in respective order of increasing free energy. **B.** Folding kinetics of CSIR 2.19 77-mer from the open chain structure (computed by the treekin program of the Vienna RNA Barriers). The plot shows the population density of meta-stable states along a simulated folding path beginning from the open chain (or the structure with the highest free energy value, in this case structure 50, black line), and eventually reaching the MFE conformation. The population density of the initial state (structure 50) quickly declines, reaching 50% at time step 1e+ 00, and later 0% before 1e+03. The global minimum (MFE—structure 1) dominates at the beginning of the refolding process, reaching over 40 % density by time 1e+02. This is followed by 5, 6, 3, 4, and later 2.

3.8 Validation of chemically synthesised truncated aptamers by non-denaturing PAGE

The sequences of the truncated derivatives of CSIR 2.11 and 2.19 were rationally designed based on secondary structure simulations and analyses. The sequences of these aptamers, along with those of the full length parents were sent to IDT to be chemically synthesised on a large scale. The chemically synthesised aptamers were 5′-biotin-modified to enable detection in the Enzyme Linked Oligonucleotide Assay (ELONA). The purity and apparent size of the aptamers were analysed by non-denaturing PAGE analysis (Figure 3-27).

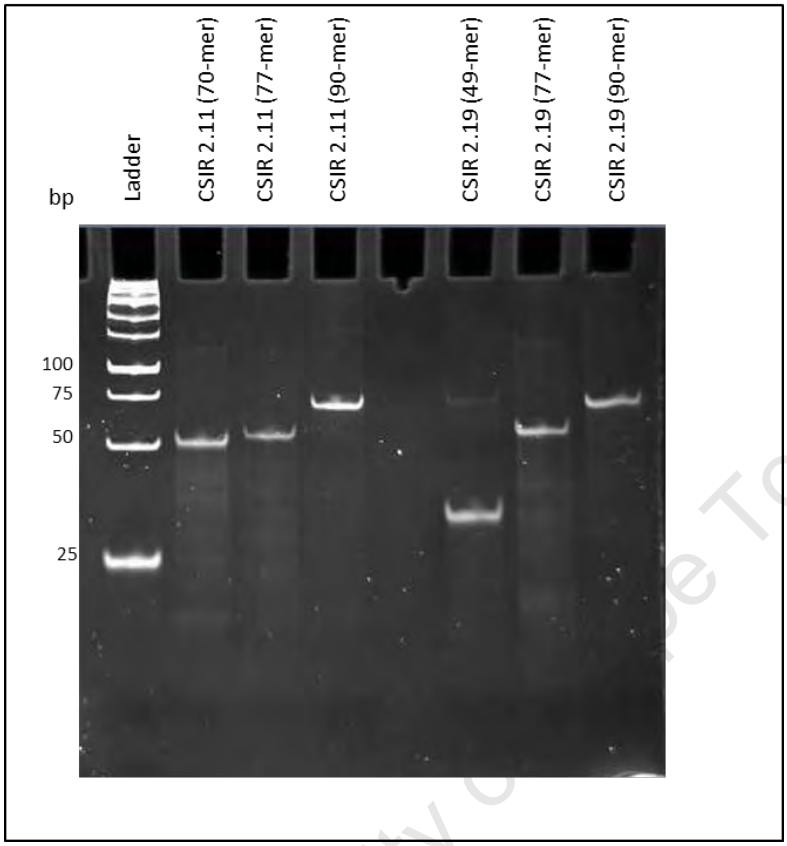


Figure 3-27: Non-denaturing PAGE analysis comparing the mobilities of the full length and truncated aptamers. Full length ssDNA aptamers CSIR 2.11 and 2.19 are compared to their respective truncations of varying length. The lanes were loaded from the smallest to the largest sequence for each aptamer. Lanes 2 to 4 contain the 70-mer, 77-mer, 90-mer of CSIR 2.11 and lanes 6 to 7- contain the 49-mer, 77-mer, and the 90-mer of CSIR 2.19 respectively. The mobilities for the truncations and full lengths are as expected, (showing decreasing mobility with increase in size). The aptamers are at around the expected sizes; between 75 and 100bp for the full length aptamers, and the truncated derivatives run slightly lower than the full length sequences. It is important to note that the ladder used is double stranded and thus does not allow for accurate size determination for single stranded DNA.

3.9 Binding of the truncated aptamers compared to the full length

To determine the binding of the truncated aptamers in comparison with that of their full length parent sequences, ELONAs were performed. The expectation here was two-fold. Firstly that having eliminated non-essential bases from the aptamer sequences, the truncated derivatives would bind the same as the full length aptamers. Alternatively, if essential bases, which were not predicted as essential by mfold, were eliminated, the truncated versions would show inferior binding to that of the full length aptamers. Secondly, the outcome would also provide experimental data to support the inferences made from the folding dynamics analyses, which predicted that the truncated aptamers would likely make for more favourable binders than the full length aptamers.

To answer these questions, ELONAs were set up with both the full length and the truncated aptamers and their relevant controls under the same conditions. The data presented herein for each aptamer are from two independent experiments, with all reactions performed in triplicate. The data was normalised to the full length aptamer plus the CFP-10 monomer absorbance, i.e. the mean relative binding of the full length (90-mer) + CFP-10 was set to 1. Comparative analyses of the relative binding values of the full length and truncated aptamers are shown for aptamers CSIR 2.11 and 2.19, respectively (Figures 3-28 and 3-29). As expected, the full length (90-mer) + CFP-10 show significantly higher relative binding than the CFP-10 alone controls (* P-value <0.001). The signal for the truncated aptamers+ CFP-10 was significantly higher compared to the CFP-10 only control (* P-value <0.001). This indicates that the truncated aptamers retained their binding functionality. Similarly, the truncated aptamers were able to detect the ESAT-6.CFP-10 heterodimer (Appendix, Figures A-2 and A-3). When comparing the relative binding of the truncated aptamers with that of the full length aptamers, no significant differences (* P-value >0.05) were observed.

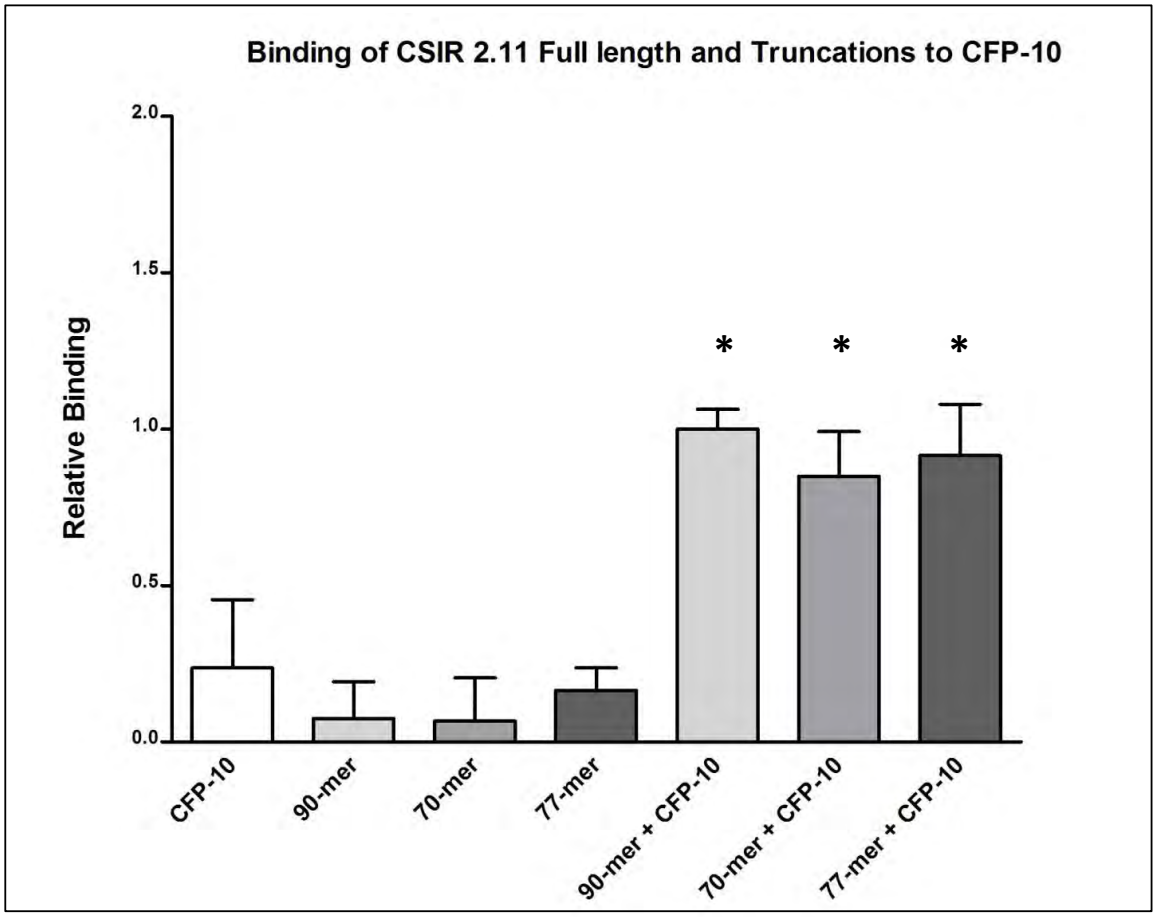


Figure 3-28: Relative binding of CSIR 2.11 full length and truncated aptamers to CFP-10 assessed by ELONA. Relative binding of the biotin-labelled aptamers to the target proteins was measured by streptavidin-HRP-catalysed reactions, detected with TMB substrate, and read as absorbance values at 450nm. The binding of the truncated aptamers is represented relative to that of the full length. Each data point represents combined triplicate values from two independent experiments; normalised against the 90-mer + CFP-10 data points. The error bars represent the standard deviation. Both the full length aptamer+CFP-10 and truncated aptamer+CFP-10 combinations were significantly higher (* P-value <0.001) than the CFP-10 alone and aptamer alone (no protein) controls. The truncated aptamers bound to the CFP-10 monomer with affinity values comparable to those of the original 90-mer aptamer, without significant differences. The truncated aptamers bound to the CFP-10 monomer with absorbance values comparable to those of the original 90-mer aptamer, without significant differences.

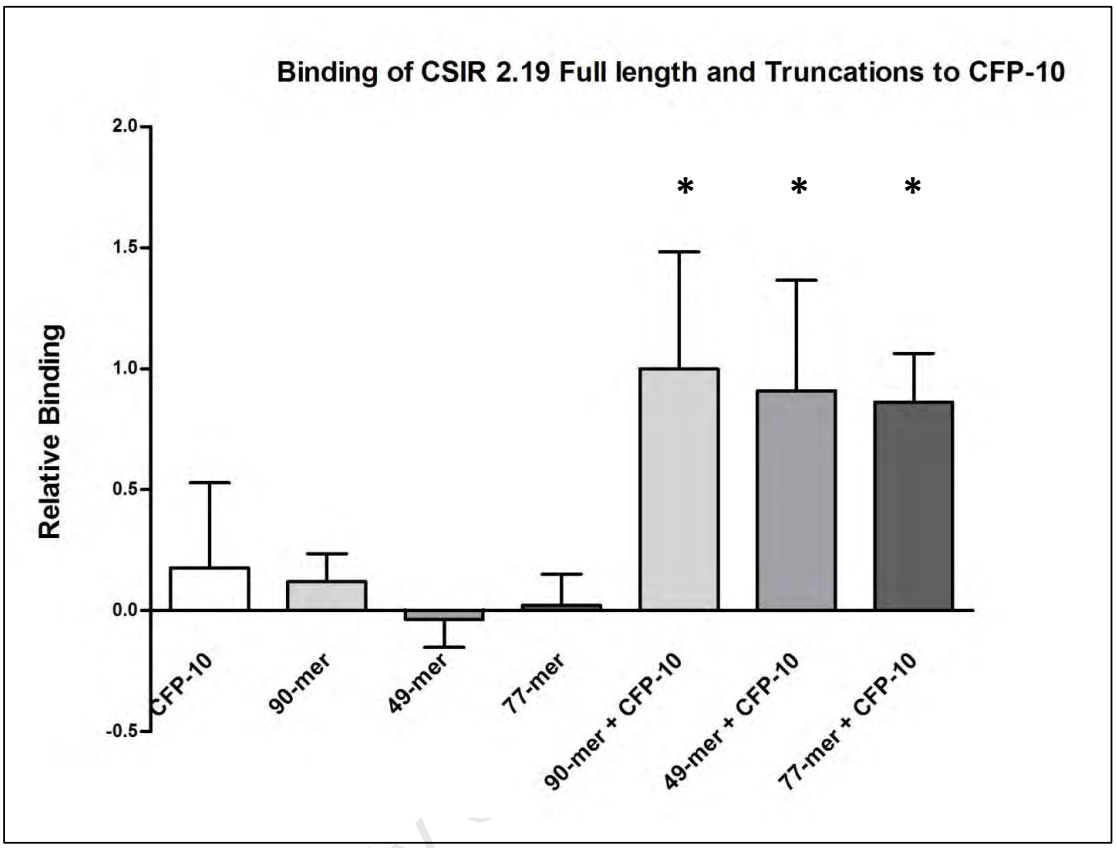


Figure 3-29: Relative binding of CSIR 2.19 full length and truncated aptamers to CFP-10 assessed by ELONA. Relative binding of the biotin- labelled aptamers to the target proteins was measured by streptavidin-HRP-catalysed reactions, detected with TMB substrate, and read as absorbance values at 450nm. The binding of the truncated aptamers is represented relative to that of the full length. Each data point represents combined triplicate values from two independent experiments; each normalised against the 90-mer + CFP-10 data points. The error bars represent the standard deviation. Both the full length aptamer+CFP-10 and truncated aptamer+CFP-10 combinations were significantly higher (* P-value <0.001) than the CFP-10 alone and aptamer alone (no protein) controls. The truncated aptamers bound to the CFP-10 monomer with affinity values comparable to those of the original 90-mer aptamer, without significant differences. The truncated aptamers bound to the CFP-10 monomer with absorbance values comparable to those of the original 90-mer aptamer, without significant differences.

CHAPTER 4

DISCUSSION

To recapitulate, this study investigated the structural and the functional properties of six anti-ESAT-6.CFP-10 aptamers, which were previously isolated by our research group and recently published (Rotherham et al 2012). Furthermore, this study explored the post-SELEX optimisation of two of these aptamers, namely CSIR 2.11 and 2.19. The binding affinities of six full length aptamers to the CFP-10 monomer were examined by SPR technology. Five of the six aptamers showed tight binding to CFP-10 with K_D values as low as 1.6 nM. One of the aptamers, CSIR 2.9, did not show concentration-dependent binding and its kinetics could not be derived using the SPR technology. EMSA was used to determine the kinetic interactions between CSIR 2.11 and the ESAT-6.CFP-10 heterodimer, however the results thereof were inconsistent and thus kinetic parameters could not be estimated. The structural properties of the aptamers were predicted by theoretical computational methods, using Zucker's mfold and the Vienna RNA package programs. These programs allowed for the prediction of secondary structure traits of the aptamers, such as the MFE structure and associated Gibbs free energy (ΔG), as well as folding energy landscapes. Predicted secondary structure features were used to forecast the target-binding motifs of the aptamers, and inform the post-SELEX optimisation of two of the aptamers by rational truncation. Fittingly, full length aptamer sequences were truncated to shorter derivatives which retained the predicted target-binding parts. The binding affinity of truncated aptamers was compared to that of their full length parent sequences through ELONA. The truncated aptamers showed comparable binding to the full length aptamers with no statistically significant differences.

DISCUSSION

4.1 Expression and purification of ESAT-6 and CFP-10

Expression and purification of histidine-tagged ESAT-6 and CFP-10 were successfully achieved, as confirmed by SDS-PAGE and immunoblotting analysis. During expression, ESAT-6 is known to enter inclusion bodies when induced at 37°C (Renshaw et al., 2002). Therefore, to help increase its solubility, the ESAT-6 bacterial cultures were grown at 25°C (Chaudhary et al., 2005, Kulshrestha et al., 2005). This allowed for the localisation of ESAT-6 in the soluble fraction rather than in inclusion bodies (Rotherham et al., 2012). Furthermore, the use of a chaotropic agent, such as urea, in the purification process helped improve the solubility of ESAT-6 (Rotherham et al., 2012).

Purification of ESAT-6 and CFP-10 was achieved using IMAC; a widely used method which involves capturing proteins containing peptide affinity tags on an immobilised metal-based matrix (Bornhorst and Falke, 2000, Ward and Swiatek, 2009). The technique is based on the interactions between a transition metal ion, in our case, Nickel (Ni^{2+}) and the amino acid side chains on the Histidine tag (Bornhorst and Falke, 2000). This method allowed for rapid protein purification and is known to potentially achieve purities of up to 95% in high yield (Bornhorst and Falke, 2000, Ward and Swiatek, 2009). However, the yields obtained in the current study were very low, in comparison to previous studies using similar protocols (Bornhorst and Falke, 2000, Renshaw et al., 2002, Ward and Swiatek, 2009). The low yields could be attributed to a number of technical factors during the expression and purification process. These unfavourable factors may have resulted from the culturing conditions, the protein extraction process (Ward and Swiatek, 2009), the purification method (Sivashanmugam et al., 2009), or a combination of these.

Although some great advances have been made in the art of protein expression over the years, the production of functional eukaryotic proteins in *E.coli* at high levels still has some challenges (Sivashanmugam et al., 2009). These challenges range from the structural properties of the gene sequence to be expressed, the stability and translational efficiency of mRNA, to the ease of protein folding and degradation of the protein by host

cell proteases (Sivashanmugam et al., 2009). To address some of these, a number of solutions can be tailored for the particular protein to be expressed. One approach could be optimising growth conditions, which are mostly target-dependent and cannot for the most part be universalised (Sivashanmugam et al., 2009). Flaws in the culturing conditions such as limited oxygen availability and increased carbon dioxide, which affects the pH of the growth medium, and plasmid loss, can adversely affect the protein expression process (Sivashanmugam et al., 2009). These problems often cause low and in some case no protein production (Sivashanmugam et al., 2009). To circumvent some of these, culture growth temperatures can be changed, as was done for ESAT-6 in previous studies (Renshaw et al., 2002). Also, instead of inducing protein expression at OD600 of 0.6, induction could be attempted at an intermediate value such as in the middle of the growth phase, to ensure high-level protein expression (Sivashanmugam et al., 2009).

For purification purposes, while the IMAC method is generally successful (Ward and Swiatek, 2009), it is reported that the method may fall short, particularly in the case of tagged proteins which are expressed at low levels or for the preparation of highly homogeneous protein samples (Bornhorst and Falke, 2000). Moreover, purity is also a concern. The non-specific binding of proteins to the IMAC-based columns is cited a major disadvantage, more so when the tagged protein is not expressed at high levels (Bornhorst and Falke, 2000). This suggests that there would likely be other proteins in the "purified protein" sample which may be included in the quantitation process, thus giving an overestimate of the yield of the protein of interest. These overestimations could affect the outcome of downstream functional assays. However, this limitation can be avoided by using the same batch of protein whose estimated concentration will be the assumed "standard" and having all other concentration estimations be derived from that standard. For the proteins expressed in the current study, however, controlling for batch-to-batch variation would have been tricky as the yields from each expression were much lower than expected.

Therefore, to circumvent the factors discussed above, and to avoid the high cost implications of reagents, consumables and time that came with the required multiple expression and purification repeats, we opted to purchase large-scale produced, HPLC-purified protein for use in further experimentation and functional assays. The purchased ESAT-6 and CFP-10 were used for the *in vitro* production of the ESAT-6.CFP-10 heterodimer complex. The formation of the heterodimer was confirmed by native PAGE and showed similar migration patterns to those previously observed by (Rotherham et al., 2012).

4.2 In-house production of ssDNA aptamers

Plasmid DNA was successfully prepared and purified, and subsequently used in PCR amplification assays to produce dsDNA using previously described protocols (Avci-Adali et al., 2009). Optimisation steps were carried out in order to maximize the yield and purity of the PCR product; and the conditions outlined in the methods section were found to be optimal. Lambda exonuclease-digestion of dsDNA was found to be effective in generating ssDNA. The lambda exonuclease enzyme selectively digests the 5′-phosphorylated strand of dsDNA, leaving the complementary single strand intact for use in functional assays (Avci-Adali et al., 2009), This method is known to be among the most effective for this purpose, when compared to other techniques such as denaturing urea-polyacrylamide gel, asymmetric PCR, and streptavidin-coated beads-biotin separation (Avci-Adali et al., 2009). For instance, the product from the asymmetric PCR has to be purified by nondenaturing electrophoresis and eluted from the gel (Avci-Adali et al., 2009). This process is not only time consuming, but results in low yields (Marimuthu et al., 2012). Furthermore, the alkaline conditions streptavidin-coated beads-biotin separation method may lead to loss of the tertiary structure of the ssDNA and consequently compromise the target-binding function of the aptamer. While exonuclease digestion provided a rapid way of producing ssDNA, the maximum yields obtained in the current study were about 30% lower in comparison to previous studies gel (Avci-Adali et al., 2009). In order to generate sufficient ssDNA for use in the aptamer characterisation

assays, high amounts of ssDNA were required, which meant repeated steps of PCR and exonuclease digestions would have to be conducted. This would be time-consuming and costly. Thus, large-scale chemical synthesis of the aptamers was opted for instead.

4.3 Characterisation of full length aptamers

4.3.1 Binding Kinetics of full length aptamers as determined by SPR

The molecular interactions between full length aptamers and the CFP-10 monomer were characterised by SPR as previously described (Joss et al., 1998, Kimple et al., 2010). The aptamers in this study showed high affinity to CFP-10 and generally displayed concentration-dependent binding. It is important to note that these binding properties refer to immobilised-ligand binding properties, which may be different from those in free-ligand assays, depending on the analyte-ligand pair in question (Win et al., 2006). Also, the coupling method used to immobilise the ligand can influence the outcome of the experiment (Joss et al., 1998, Kimple et al., 2010). Random direct coupling methods such as amine coupling the ligand to the dextran surface of the chip, which was used in this study, may introduce surface-imposed heterogeneity by reacting with residues in or near the recognition site (Joss et al., 1998, Kimple et al., 2010). This effect may distort the binding properties of the ligand, and can be minimised by capturing the ligand using antibodies, or streptavidin–biotin interactions (Joss et al., 1998, Kimple et al., 2010). However, the amine coupling of CFP-10 to the SPR sensor chip did not seem to adversely affect binding to the aptamers in this study.

The kinetics data obtained in the current study indicate tight binding of the aptamers to CFP-10, with K_D values in the nanomolar ranges, which are within the expected range for aptamers and consistent with earlier findings by others (Chiu, 2009, Famulok et al., 2007, Pendergrast et al., 2005, Proske et al., 2005). Of the six aptamers studied, CSIR 2.19 showed the best binding kinetics, with the lowest K_D of 1.6 nM. The low K_D values are an indication of high affinity to the target and support the feasibility of using these aptamers in downstream clinical applications (Stoltenburg et al., 2007). Interestingly,

CSIR 2.11, which previously showed the highest absorbance reading using ELONA (Rotherham et al., 2012), did not exhibit the lowest K_D when SPR was used. This discrepancy could be attributed to the technical sophistication of the SPR technology compared to that of the ELONA. SPR assays give better concentration-dependent responses and have better sensitivity than ELISA (Lee et al., 2008). This is primarily because the SPR instrument records binding events in real time in contrast to the end-point binding result in the ELONA (Luzi et al., 2003, Nguyen et al., 2009).

Kinetic data is generally interpreted according to the model of analysis used, and the choice of model is dependent on the nature of the experiment. The kinetic parameters of the aptamers evaluated in this study were obtained using the Langmuir model assuming 1:1 binding. This model is the simplest and most widely used model (Gopinath, 2010, Joss et al., 1998). Although the model is not considered a precise or perfect representation of what happens on the chip surface (Gopinath, 2010, Joss et al., 1998) , using this model to perform kinetic ranking of different analytes binding to the same ligand but from data which do not represent homogenous binding is considered justifiable (Gopinath, 2010, Joss et al., 1998). Therefore, the apparent dissociation constants determined this way are sufficient for most purposes (Gopinath, 2010, Joss et al., 1998). For kinetic evaluations in the current study, the results were found to be within acceptable ranges of the relevant statistical measures. One such statistical measure is the Chi^2 value, which measures the "closeness or goodness of fit" of experimental data to the theoretical model used to analyse data (Gopinath, 2010, Joss et al., 1998, Tombelli et al., 2005, Win et al., 2006). It is recommended that for good fitting data, the Chi^2 value should be of the same order of magnitude as the noise signal (Biacore, 1997). The Chi^2 values obtained in the current study were in the lower single digits while the noise signal was in the lower double digits; on average, no more than 20 RU. To further support the reliability of the kinetics data obtained in this study, the K_D values ought to be within the expected range for the type of molecular interaction being assessed (Li et al., 2006, Luzi et al., 2003, Potyrailo et al., 1998, Win et al., 2006). For aptamer-protein interactions in

particular, these values are generally within lower nanomolar ranges (Li et al., 2006, Luzi et al., 2003, Potyrailo et al., 1998, Tombelli et al., 2005, Win et al., 2006). Thus, the kinetic data obtained in this study are considered satisfactory.

On the other hand, CSIR 2.9 neither displayed concentration-dependent binding nor kinetic parameters that fit the Langmuir 1:1 model used for analysis. SPR-based kinetic evaluations can be affected by several factors including the purity of reagents, the immobilisation procedure and level, ligand activity, flow rate and analyte concentration range (Biacore, 1997, Gopinath, 2010, Joss et al., 1998). The problem with respect to the CSIR 2.9 kinetics could not be attributed to the activity of the ligand (CFP-10), as the aptamers that were injected over the same ligand surface following aptamer CSIR 2.9 and subsequent regeneration still showed concentration-dependent binding and good kinetic parameters. It could be that the conditions in which the binding reactions were taking place, which suited the other five aptamers, may not have been favourable for CSIR 2.9. Alternatively, considering that the SPR is a more sensitive (Lee et al., 2008) and possibly more accurate (Biacore, 1997, Gopinath, 2010, Joss et al., 1998) technique than the ELONA, the CSIR 2.9 aptamer may not be a high affinity binder as was previously suggested by the initial ELONA screening assays performed by (Rotherham et al., 2012).

Altogether, characterisation and quantification of the interactions between the anti-ESAT-6.CFP-10 aptamers and the CFP-10 protein were achieved using SPR technology in the current study. The aptamers showed K_D values in the nanomolar range, indicating high affinity to the CFP-10 protein. Accordingly, the five aptamers whose kinetics were determined in the current study, particularly CSIR 2.19, are good candidates for use as detection probes for CFP-10 proteins secreted by MTB. Although, including controls such as an unrelated aptamer sequence and assessing binding to CFP-10, and immobilising an unrelated protein to the sensor chip was necessary in order to conclusively confirm aptamer selectivity to the target proteins (Dupont et al., 2010, Lee et al., 2008).

4.3.2 Evaluation of Binding affinity of aptamer CSIR 2.11 by EMSA

EMSA is another technique that can be used to characterise molecular interactions between nucleic acids ligands and cognate proteins. To do this, the electrophoretic mobility of nucleic acid-protein complexes is measured relative to that of free (unbound) nucleic acids (Hellman and Fried, 2007, Demarse et al., 2009). The method is based on the premise that nucleic acid-protein complexes migrate slower than the corresponding free nucleic acids. The current study employed the EMSA approach with a view to examine the binding of CSIR 2.11 to the ESAT-6.CFP-10 dimer and to the CFP-10 monomer, respectively. CSIR 2.11 was successfully labelled with radioactive nucleotides to enable the visualisation and quantification of aptamer-protein interactions. However, the results were inconclusive.

When lower protein concentrations (250 to 1000 nM) were used, two rows of bands were observed in the aptamer-protein lanes, suggesting modest mobility shift. Furthermore, the CSIR 2.11-CFP-10 complexes migrated less than the CSIR 2.11-ESAT-6.CFP-10 dimer complexes, in the same way that the CFP-10 monomer migrated less than the dimer under non-denaturing PAGE conditions as previously reported (Rotherham et al., 2012). Typically, nucleic acid mobility and the amount of free or unbound nucleic acid usually decrease with an increase in the concentration of the cognate protein (Hellman and Fried, 2007). However, this trend was not apparent in the current study when using the lower protein concentration range. In contrast, when higher protein concentrations (1 μM to 10 μM) were used, only single bands were observed in the aptamer-protein lanes, and the shift was in the opposite direction to what was expected. The CSIR 2.11-ESAT-6.CFP-10 dimer complexes appeared to have migrated further than the CSIR 2.11 only control. This outcome cannot be pinned down to a single factor as the results varied considerably across experimental repeats.

Nonetheless, an EMSA shift is primarily influenced by three key factors: molecular weight, charge adjustment, and changes in the topology of the migrating nucleic acid

(Alves and Cunha, 2012, Hellman and Fried, 2007). During electrophoresis, the negatively charged nucleic acid migrates towards the anode. The extent of this migration is proportional to the molecular weight of the nucleic acid (Alves and Cunha, 2012, Hellman and Fried, 2007), which is related to nucleic acid chain length (Hellman and Fried, 2007). In addition, the net charge of the nucleic acid-protein complex, which is also affected by the protein charge, plays a role in mobility (Alves and Cunha, 2012, Hellman and Fried, 2007). If the protein charge is highly negative (as would be the case when its isoelectric point is lower than the pH of the electrophoresis buffer), the protein is likely to move towards the anode and may cause the nucleic acid-protein complexes to migrate very close to or further than the nucleic acid (Hellman and Fried, 2007). In terms of nucleic acid topology, if the binding of a protein does not induce any considerable bend on the nucleic acid, the conformational contribution of that protein to mobility retardation would be minimal (Hellman and Fried, 2007). These factors may have independently or collectively contributed to the atypical results observed in the current study. Interestingly, the binding of a protein to a nucleic acid has been previously shown to accelerate rather than hinder mobility (Hellman and Fried, 2007). This outcome was observed in a study by (Sandman et al., 1990), where the protein induced nucleic acid condensation of a relatively large linear DNA. In this case, the conformational change of the DNA was a stronger factor than the overall molecular weight increase, which in turn caused acceleration rather than a decrease in DNA mobility (Hellman and Fried, 2007).

Although EMSA is considered a relatively fast, effective and highly sensitive technique which can be used to characterise molecular interactions (Chorley et al., 2008, Hellman and Fried, 2007); it has limitations. Apart from the samples not being at chemical equilibrium during electrophoresis (Hellman and Fried, 2007), the mobility of nucleic acid-protein complexes is influenced by several factors other than the size and concentration of the protein (Chorley et al., 2008, Hellman and Fried, 2007), such as those discussed above. These factors introduce multiple variables that influence the outcome of the assay. Moreover, optimising EMSA is generally a matter of trial and error

(Alves and Cunha, 2012, Demarse et al., 2009, Hellman and Fried, 2007), which can be time-consuming and challenging (Chorley et al., 2008). Having employed the recommended electrophoretic conditions such as gel type, buffer conditions, and complex stabilising factors (Hellman and Fried, 2007, Laniel et al., 2001), the EMSA protocol in the current study could not be optimised. Consequently, the strength of the aptamer-target interactions and associated kinetic parameters could not be determined using this method

4.4 Structural features of the aptamers and their predicted contribution to function

Secondary structures of the anti-ESAT-6.CFP-10 aptamers were predicted using theoretical algorithms. Zucker's mfold was used to estimate the MFE for each aptamer. These structures provided a geometric and thermodynamic framework for potential tertiary structures (Flamm et al., 2004). In nucleic acid folding, double stranded helices typically form before the tertiary contacts of nucleic acids complete the formation of the three-dimensional structure (Flamm et al., 2004). Furthermore, the major portion of the free energy of nucleic acid molecules is accounted for by secondary structures (Flamm et al., 2004, Heine et al., 2006, Wolfinger et al., 2004, Wuchty et al., 1998). Thus, the predicted secondary structures and associated thermodynamics presented in the current study provided a theoretical model of the potential folding patterns of the aptamers that confer structural compatibility for target-binding.

4.5 Optimisation of aptamers through rational truncation

Aptamer functionality can be improved by the post-SELEX optimisation through the truncation of the aptamer sequences to their minimal target-binding motifs (Savory et al., 2010, Shangguan and Tang Z., 2007, Zhou, 2010). Two candidate aptamers CSIR 2.11 and CSIR 2.19 were chosen for post-SELEX optimisation in the current study. CSIR 2.11 was chosen because it was better characterised and showed potential clinical utility when tested on clinical sputum samples (Rotherham et al., 2012), while CSIR 2.19 was

selected based on its lowest dissociation constant value relative to the other aptamers. The truncation of aptamers CSIR 2.11 and 2.19 was conducted on the basis of their predicted secondary structures. Nucleic acid secondary structures have been widely used as a "surrogate for function" (Flamm et al., 2004, Mathews and Turner, 2006, Nonaka et al., 2010, Wolfinger et al., 2004, Wuchty et al., 1998). This notion is further demonstrated by *in silico* selections (Patzel 2004; Hall et al. 2007), and the post-selection analyses of aptamer secondary structures (Matthews et al 2008).

For the two aptamers chosen for truncation in this study, the objective was to remove non-essential nucleotides on the aptamer sequences and retain the parts of the sequences that are essential for target binding. This would potentially give the aptamers better structural stability and consequently improved functionality and specificity (Nonaka et al., 2010, Potty, 2008), and reduce their dimensions to make them compatible with downstream biosensor design (Zhou, 2010). Moreover, the design of shorter-sequence aptamers would not only make them easier to produce on a larger scale, but also more cost effective (Shangguan and Tang Z., 2007). To achieve this objective, the regions of the aptamer sequences that contribute to target-recognition and binding were deduced based on the mfold-predicted MFE structures. Based on the MFE structures, the sequences of four truncated derivatives (two per aptamer) were designed. This was done by deleting several nucleotides in either direction of the primer regions of full length aptamer sequences.

The folding of the truncated sequences was subsequently simulated, and the truncated derivatives retained the secondary structures and free energies of the parent sequences. The working hypothesis in this case was that maintaining the overall structure of the predicted target-binding region of the aptamer would be essential for retaining the functional activity of the truncated versions. The removal of nucleotides from the primer sequences of the aptamers did not affect aptamer folding, and would therefore not likely affect their binding. This finding is in agreement with previously published work, which

demonstrated through secondary structural analysis of over 2000 aptamers, that the binding of aptamers is not generally enhanced or constrained by constant primer regions (Cowperthwaite and Ellington, 2008).

The current study employed a bioinformatics approach to examine the influence of sequence length and secondary structures on aptamer function. The secondary-structure guided, rational truncation method utilised in the current study, was shown in previous studies to be reliable (Potty, 2008, Shangguan and Tang Z., 2007). The method provided a rapid and effective approach through which to shorten the aptamer sequences, and eliminated the laborious technical demands that come with the traditional methods such as enzymatic footprinting or partial hydrolysis (Bing et al., 2010, Nonaka et al., 2010).

4.6 Folding energy landscapes of full length and truncated ESAT-6.CFP-10 aptamers

Secondary structure prediction methods generally identify functional structures of nucleic acid molecules on the basis of thermodynamic equilibrium, without factoring in the dynamics of the folding process (Washietl et al., 2012). Therefore, to get a broader picture of the folding dynamics of the aptamers used in the current study, and to determine whether there would be any notable differences between the full length and truncated aptamers in this regard, folding energy landscape analyses were performed using the computational methods.

Thermodynamically, a folding energy landscape provides a visual depiction of how close the MFE is to its thermodynamic neighbours and how much energy would be required for a molecule to switch between conformations (Wolfinger et al., 2004, Wuchty et al., 1998). The greater the energy difference (barrier height) between the MFE and its thermodynamic neighbours (suboptimal structures), the less likely it is that these suboptimal structures will exist—or be reached within a given folding time (Heine et al., 2006). Another aspect of folding landscapes is the time factor, which speaks to the

kinetics of the folding process. The folding of an aptamer into a target-cognate conformation is a time-sensitive process (Flamm et al., 2000, Flamm et al., 2002).

Folding kinetics would predict how long it would take for the MFE of a particular aptamer to reach maximum potential occupancy within a given population of possible conformations (Flamm et al., 2000, Slutsky and Mirny, 2004, Wuchty et al., 1998). In addition, the folding kinetics gives an indication of whether the folding process is likely to be affected by "kinetic traps", i.e. the molecule would be stuck in alternative suboptimal conformations along the path to the MFE (Flamm et al., 2000, Heine et al., 2006, Wolfinger et al., 2004).

To answer some of the questions outlined above for CSIR 2.11 and 2.19 and their respective truncated versions, energy landscape analyses were conducted by computing barrier trees using the Vienna RNA Package *Barriers* program. The algorithm yielded barrier trees of the low energy portion of the folding landscape of the aptamers. The program *treekin* (part of *Barriers*) was used to compute the folding dynamics of the DNA molecules on the basis of a precomputed barrier tree. Using the discrete model of secondary structure folding provided by the program, approximate descriptions of the folding dynamics of the selected aptamers at logarithmic time-scales were presented. These estimations provided a framework within which to predict whether aptamer sequences are likely to switch between conformations, and at which time scales (Flamm et al., 2000, Heine et al., 2006, Wolfinger et al., 2004).

For the simulations in this study, based on the nature of the computational tools used (Flamm et al., 2000, Heine et al., 2006, Wolfinger et al., 2004), the focus was not on the time the system would take to reach a thermodynamically stable equilibrium distribution, but rather the refolding time from a selected start structure (representing the denatured, open-chain structure) to the ground state global minimum or the MFE. The outcome of the computational analyses would give an idea of whether or not the proposed "search and slide" process of the protein to find its cognate site on the aptamer sequences

(Hermann and Patel, 2000) would occur fast enough for the more thermodynamically stable structure (MFE) to be favoured, without the aptamer sequence being trapped into some other conformation that does not provide the best affinity for the protein.

The barriers data for the 90-mer CSIR 2.11 revealed that the MFE reached its maximum occupancy later in the folding time scale, and did not turn out to be in the "kinetically preferred" subtree with more stable conformations. In addition, the barrier height between this subtree and the one containing the MFE is not high, which suggests ease of transition between the two basins (Heine et al., 2006). This is also supported by the percentage occupancy values on the folding kinetics graphs. The folding kinetics of a barrier tree determine whether the molecule has sufficient time to switch into the optimal subtree; the one representing a class of more thermodynamically stable structures, or whether it will remain trapped in the suboptimal subtree (Heine et al., 2006). Based on the data obtained for the 90-mer CSIR 2.11, it is plausible to predict that the suboptimal structures represented by local minima 4, 8, and 44 will be present in considerable amounts in the structural populations of aptamers leading to the MFE. The data also forecast that while the MFE structure of the 90-mer CSIR 2.11 that occurs in solution may be fairly stable once formed, the molecules are likely to flip-flop between different conformations particularly those in the basin containing local minima 4 and 8. The 70-mer showed better folding kinetics than the 90-mer and 77-mer, with the MFE being the most likely to occur within a biologically reasonable timeframe, and the more thermodynamically preferred.

The 90-mer CSIR 2.19, showed better folding kinetics than did the 90-mer CSIR 2.11 in that the subtree that is the most occupied and the more likely kinetically preferred is that which contains the MFE. However, the MFE is likely to be strongly challenged by local minimum 2. This data corroborates the MFE estimations by mfold, whose output was three MFE structures with almost equal ΔG. While this implies a likely conformation switch between the two structures, the percentage occupancy of the MFE within the set

time-scale suggests that the MFE will win the challenge and form long before structure 2 does. The 49-mer CSIR 2.19 aptamer showed similar folding kinetics to that of the 90-mer, but with a higher MFE population of almost double that of the 90-mer. The 77-mer CSIR 2.19 showed better kinetics than that of the 90-mer and 49-mer. Also, the MFE and local minimum 2 are found in different basins separated by a high barrier, suggesting that a conformational switch is unlikely.

Overall, there were some notable differences in the folding dynamics between the full length and truncated aptamers. The general trend was that the MFE was reached earlier in the truncated than in the full length aptamers. These results indicate that the truncated aptamers refold into the MFE population faster than the full length aptamers. In addition, the MFE's of the truncated versions had higher population occupancy in comparison to those of the full length aptamers. Furthermore, the number of "basins of attraction" seemed to decrease in proportion with the sequence length. Taken together, the truncated aptamers displayed better folding dynamics than their full length parent sequences. These findings support the notion that the post-SELEX optimisation of aptamer sequences by truncation can potentially confer the aptamers with better structural stability (Savory et al., 2010, Shangguan and Tang Z., 2007, Zhou, 2010).

It is however worth noting that the folding kinetics simulations presented herein were performed within a designated time-frame in which the folding trajectories do not show the rest of the local minima (which along with those represented in the plots will collectively add up to a total probability value of 1), and that these "unrepresented structures" may dominate the folding pathway and act as "kinetics traps", preventing the sequence from reaching the MFE structure. It is also important to note that the models used to predict secondary structures are limited as they do not account for solvation or intermolecular influences (Flamm et al., 2000, Slutsky and Mirny, 2004, Wuchty et al., 1998). Furthermore, the parameters of nucleic acid folding estimated *in silico* are performed in the absence of the target (Gao et al., 2006). Therefore, while these

theoretical models are able to provide useful information about the functional properties of nucleic acids, their outputs should not be considered an exact representation of what occurs in the natural conditions of the nucleic acid.

4.7 Functional characterisation of truncated aptamers

4.7.1 Binding of truncated anti-ESAT-6.CFP-10 aptamers compared to full length parent sequences

From the ELONA, the ability of the truncated derivatives of CSIR 2.11 and CSIR 2.19 to recognise target proteins was comparable to that of the full-length parental aptamers. There were no significant differences between the binding of each full length aptamer and their respective truncated versions. These data demonstrate that all the truncated versions of the CSIR 2.11 and CSIR 2.19 aptamers retained the nucleobases, and consequently, the structural elements that contribute to the overall target-binding conformations of these aptamers. Furthermore, since target-recognition was retained in the truncated derivatives, it is reasonable to deduce that the deleted regions do not play a crucial role in target-binding. Additional controls for the ELONA such as an unrelated aptamer sequence and protein should be included in future to illustrate the specificity of the aptamers (Dupont et al., 2010, Lee et al., 2008). Notwithstanding, the observation that the truncated derivatives retained the binding to recombinant protein suggests that they are likely to detect TB in clinical samples.

Furthermore, the ELONA results of the truncated aptamers support the feasibility of employing computational secondary structure prediction methods in deducing the target-binding motifs of aptamers, as well as using this information to optimise the aptamers through the elimination of non-essential nucleotides (Nonaka et al., 2010, Rockey et al., 2011). This “rational truncation” approach (Rockey et al., 2011), can provide a high-throughput method for optimising aptamer structures for target-recognition (Gao et al., 2006). While the approach is limited by the reach of the computational tools used, the basis for its use is still plausible given that nucleic acid secondary structures are

considered a "surrogate for function" (Cowperthwaite and Ellington, 2008) for nucleic acids which adequately depicts aptamer-ligand interactions (Patzel, 2004).

Since the discovery of aptamers and the description of SELEX over two decades ago, a number of methods have been devised through which to exploit and improve the use of aptamers. These methods which are aimed at understanding aptamer-target interactions include delineating sequence contributions to functions and fittingly truncating sequences to shorter derivatives in order to achieve the desired dimensions for specific applications. Reduction of aptamer length has been proposed to lead to a population of molecules with more stable secondary structures containing a favourable fold that better fit the protein-binding site (Mallikaratchy et al., 2011, Nonaka et al., 2010, Potty, 2008, Wuchty et al., 1998). Even though no significant improvement in binding affinity was observed in the ELONA evaluations in the current study, the truncated versions are still preferable due to the beneficial features described in previous sections. Also, the functional improvement of the truncation may have been missed by the ELONA, and could possibly be demonstrated in a much more sensitive technique such as SPR (Lee et al., 2008) or AuNP-based sensors (Xie and Walton, 2010, Xia et al., 2010).

All things considered, I propose that the post-SELEX optimisation of the aptamers by rational truncation resulted in shorter aptamers that retained the binding functionality of the original full length aptamers. Furthermore, truncation of the full length sequences resulted in more thermodynamically and kinetically favourable structures as indicated by the folding energy landscape data. However, the folding behaviours described by these data are theoretical and do not fully represent folding patterns of the aptamers in their native environment. Nonetheless, smaller-sized aptamers allow for more desirable dimensions and are easier to manipulate in the process of designing different smaller, sensitive sensing devices (Zhou, 2010). Further truncations of the anti-ESAT-6.CFP-10 aptamer sequences may be explored. These can include simultaneous removal of nucleotide bases from both the 5' and 3' ends, and even deletions of more nucleotides in

order to generate the shortest possible functional aptamers. Another way of optimising selected high affinity aptamers such that they are more structurally stable for downstream applications would be to modify the aptamers to have more stabilised base stems, for instance by introducing more GC base pairs (cited by (Win et al., 2006). Previous studies have reported that sequences lacking strong or stabilised base stems will likely adopt a number of potential structures, while the stabilised base stem can significantly reduce this population diversity (Win et al., 2006). However, these modifications may only have minimal effects on the binding properties of the sequences (Win et al., 2006).

4.8 Proposed future studies

Proposed future studies will include testing the truncated aptamers for their ability to detect their cognate target in clinical sputum samples, and discriminate between TB and non-TB samples. Following this proof-of concept study, the clinical application of the truncated aptamers can be examined in a larger sample size as well as a variety of biological fluids used in TB diagnosis. Once optimised for detection in clinical samples using the ELISA-based method, the truncated aptamers could then be modified accordingly and integrated into a simple detection system. For instance, the aptamers could be conjugated to gold nanoparticles (AuNPs) to produce a one-step PoC TB diagnostic kit (Wang et al., 2008, Xia et al., 2010, Xu et al., 2009). The PoC diagnostic kit could be in the format of a simple probe solution (Xia et al., 2010, Xie and Walton, 2010) or lateral flow device (Liu et al., 2009, Liu et al., 2006), which would allow for the rapid, colorimetric detection of MTB antigens such as CFP-10 and ESAT-6. Thus, to optimise the design of the system, recombinant ESAT-6.CFP-10 antigens would be used as targets of detection, and the sensor could later be tested on a cohort of relevant clinical samples. To improve the specificity of the aptamer-based detection system, the anti-ESAT-6.CFP-10 aptamers could be used in pairs. Aptamer-pairing will likely multiply the specificity of the monoclonal aptamers for the intended analyte in a complex sample matrix (Gold et al., 2012). This approach would reduce the likelihood of the aptamers

being occupied by other proteins in the analyte sample, consequently reducing the noise signal and improving the detection accuracy (Gold et al., 2012).

CHAPTER 5

CONCLUSIONS

5.1 Summary of findings

This work focused on the post-SELEX optimisation and functional analysis of aptamers that bind the ESAT-6.CFP-10 MTB antigens, which are putative biomarkers of TB. To do this, the intermolecular and intramolecular properties of the aptamers were studied using experimental and bio-informatics approaches. The binding properties of the aptamers were determined using SPR technology, EMSA and ELONA. SPR was used to determine dissociation constants of the aptamer-CFP-10 monomer interactions, while EMSA was carried out to interrogate the aptamer-ESAT-6.CFP-10 interaction. Binding kinetics of the five aptamers as evaluated by SPR showed tight and concentration dependent binding to the target CFP-10 monomer, with dissociation constant values in the lower nanomolar ranges. The aptamer that showed the best binding kinetic interactions to the CFP-10 target was CSIR 2.19, with a dissociation constant (K_D) of 1.6 nM. The EMSA experiments yielded inconclusive results, and thus dissociation constants could not be derived from this data. Structural characterisation of the aptamers was performed using mfold and Vienna RNA software packages, to delineate the secondary structures and associated folding dynamics.

Based on the computationally predicted secondary structures, the aptamers were rationally truncated to remove non-essential nucleotides which are not involved in target-binding. Truncation was carried out for potential optimisation of the aptamers with downstream considerations such as improved binding functionality and specificity, and cost, prior to conjugating them to reporter systems such as nanoparticles. The post-SELEX optimisation of the aptamers by rational truncation yielded shorter and lower-cost molecules which bound to the CFP-10 target protein with affinities comparable to those of the full length sequences, and also showed improved functional stability as

predicted by theoretical models when compared to their full length parental sequences. These findings confirm that it is feasible to rationally truncate the anti-ESAT-6.CFP-10 aptamers based on predicted secondary structures, and preserve or improve their binding properties. Collectively, these findings suggest that the truncated anti-ESAT-6.CFP-10 aptamers can be used as candidate probes for the detection of the MTB antigens and consequent diagnosis of TB.

5.2 Conclusions and implications of the work

This study was undertaken as a part of the process towards developing detection probes that could potentially discriminate TB from non-TB samples. The key objectives of this study were the characterisation and post-SELEX optimisation of selected anti-ESAT-6.CFP-10 ssDNA aptamers previously isolated by our research group (Rotherham et al., 2012). The results of the current study present the binding kinetics of six anti-ESAT-6.CFP-10 aptamers, theoretical predictions of the structural contributions of these aptamers to their target-recognition function, the feasibility of post-SELEX optimisation of aptamers by rational truncation, as well as experimental evidence to support the functional performance of the rationally truncated aptamers. The rationally truncated aptamers were able to recognise their cognate target in a comparable manner to their full length parental counterparts, suggesting that the truncated aptamers will likely be able to detect TB in clinical samples. The clinical utility of the resultant optimised aptamers will, in future studies outside the scope of this thesis, be tested on TB and non-TB clinical samples in a small scale case-controlled clinical study. This will determine the potential of the truncated aptamers as candidate probes for use in the design and development of affordable; sensitive; specific; user-friendly; rapid; equipment-free and deliverable to end-users (ASSURED) PoC TB diagnostics.

APPENDIX A

A.1.1 Kinetics analysis parameters for the full length anti-ESAT-6.CFP-10 aptamers

The BIAevaluation-generated parameters used to calculate the kinetics of the full length aptamers are shown in Tables 6-1 to 6-5 below.

Table A-1: Kinetics analysis parameters for aptamer CSIR2.2

Flow cell	Dissociation time (s)	kd (1/s)	Chi2	Association time (s)	ka (1/Ms)	kD (M)	Chi2
1	400-600	1.18E-03	0.346	50-120	7.17E+04	1.65E-08	0.0497
2	400-600	1.02E-03	0.653	50-120	4.29E+04	2.39E-08	0.0426
3	400-600	9.42E-04	0.936	50-120	3.90E+04	2.41E-08	0.066
Average		**1.05E-03**			**5.12E+04**	**2.15E-08**	
Standard Deviation		**0.000121**			**17860.29**	**4.33E-09**	

Table A-2: Kinetics analysis parameters for aptamer CSIR2.15

Flow cell	Dissociation time (s)	kd (1/s)	Chi2	Association time (s)	ka (1/Ms)	kD (M)	Chi2
1	400-600	6.26E-04	0.72	50-120	5.09E+04	1.23E-08	1.09
2	400-600	6.20E-04	0.558	50-120	5.38E+04	1.15E-08	1.14
3	400-600	6.05E-04	0.561	50-120	4.11E+04	1.47E-08	1.36
Average		**6.17E-04**			**4.86E+04**	**1.28E-08**	
Standard Deviation		**1.08E-05**			**6655.073**	**1.67E-09**	

APPENDIX A

Table A-3: Kinetics analysis parameters for aptamer CSIR 2.19

Flow cell	Dissociation time (s)	kd (1/s)	Chi^2	Association time (s)	ka (1/Ms)	kD (M)	Chi^2
1	400-800	2.08E-04	1.04	50-120	2.05E+05	1.02E-09	0.332
2	400-800	1.88E-04	0.968	50-120	1.18E+05	1.59E-09	0.158
3	400-800	1.92E-04	0.895	50-120	8.77E+04	2.19E-09	0.18
Average		**1.96E-04**			**1.37E+05**	**1.60E-09**	
Standard Deviation		**1.0583E-05**			**60891.13**	**5.85E-10**	

Table A-4: Kinetics analysis parameters for aptamer CSIR 2.21

Flow cell	Dissociation time (s)	kd (1/s)	Chi2	Association time (s)	ka (1/Ms)	kD (M)	Chi2
1	400-800	7.10E-04	0.088	50-120	5.16E+04	1.38E-08	0.18
2	400-800	5.58E-04	0.112	50-120	5.79E+04	9.64E-09	0.18
3	400-800	5.31E-04	0.201	50-120	7.52E+04	7.06E-09	0.341
Average		**6.00E-04**			**6.16E+04**	**1.02E-08**	
Standard Deviation		**9.65E-05**			**12219.79**	**3.4E-09**	

Table A-5: Kinetics analysis parameters for aptamer CSIR 2.9

Flow cell	Dissociation time (s)	kd (1/s)	Chi2	Association time (s)	ka (1/Ms)	kD (M)	Chi2
1	400-800	5.06E-04	0.391	50-120	4.95E+04	1.02E-08	1.7
2	400-800	4.88E-04	0.351	50-120	5.88E-04	8.30E-09	0.188
3	400-800	5.57E-04	0.415	50-120	1.30E+04	4.93E-09	0.33
Average		**5.17E-04**			**2.08E+04**	**7.81E-09**	
Standard Deviation		**3.5791E-05**			**25662.88**	**2.67E-09**	

APPENDIX A

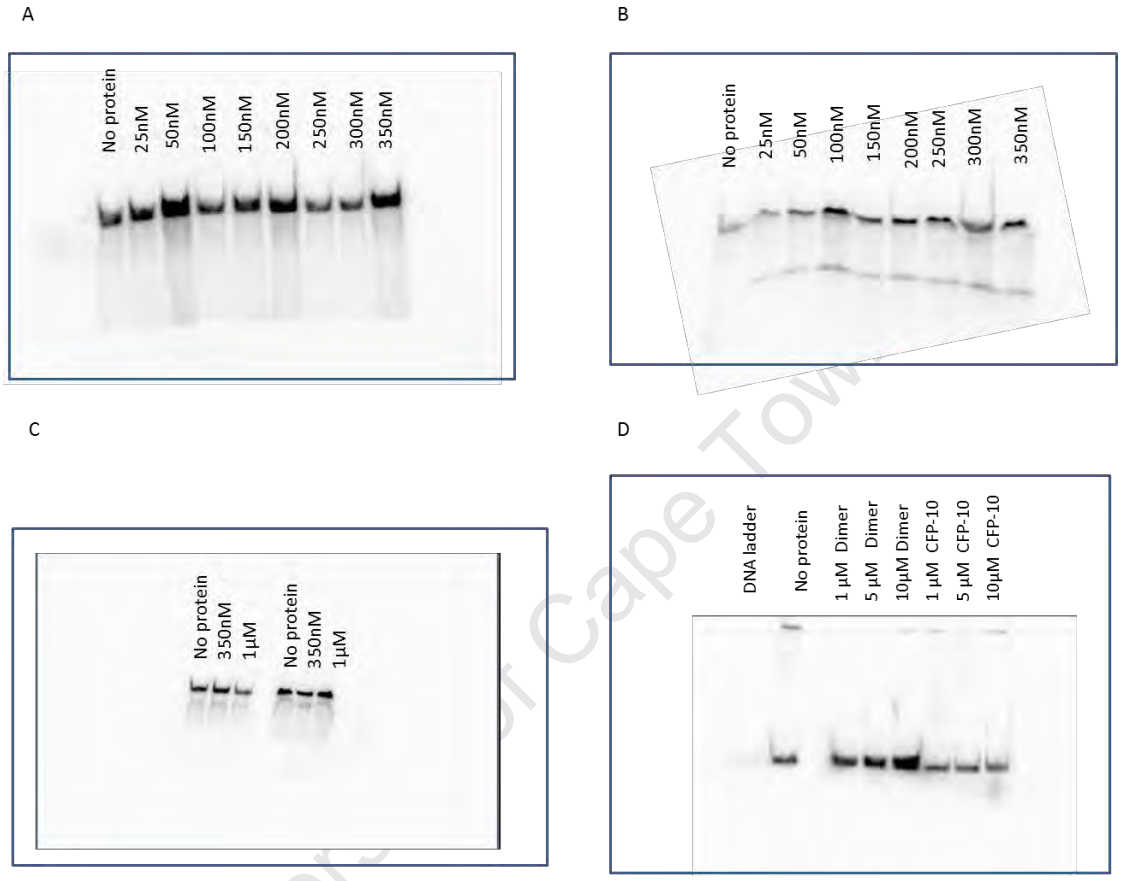


Figure A-1: Gel-shift analysis of the CSIR 2.11 aptamer. Increasing amounts of protein (either CFP-10 or ESAT-6.CFP-10 dimer-indicated above wells) were mixed with 32P-end labelled CSIR 2.11. Samples were run on a native 8% Polyacrylamide gel. The gels were dried and exposed to a phosphor screen and viewed on a phosphorimager. Two sets of protein concentrations were studied. A and B are phosphor images from initial optimisation assays. A. More concentrated protein ESAT-6.CFP-10 Dimer (10 µM) was used as starting concentration and diluted accordingly. This meant that smaller volumes of protein to use. B. a more diluted Dimer1µM was used as starting stock from which to make the dilutions, which means slightly larger volumes of protein than in gel 1, and added extra dimerization buffer (1 final v/v) to make the final binding reaction conditions the same those as in gel 1). C and D are representative gels of EMSAs that were subsequently performed using the protein dilution protocol in B, at the listed concentrations; C-used the dimer, and D; both the dimer and CFP-10 monomer

APPENDIX A

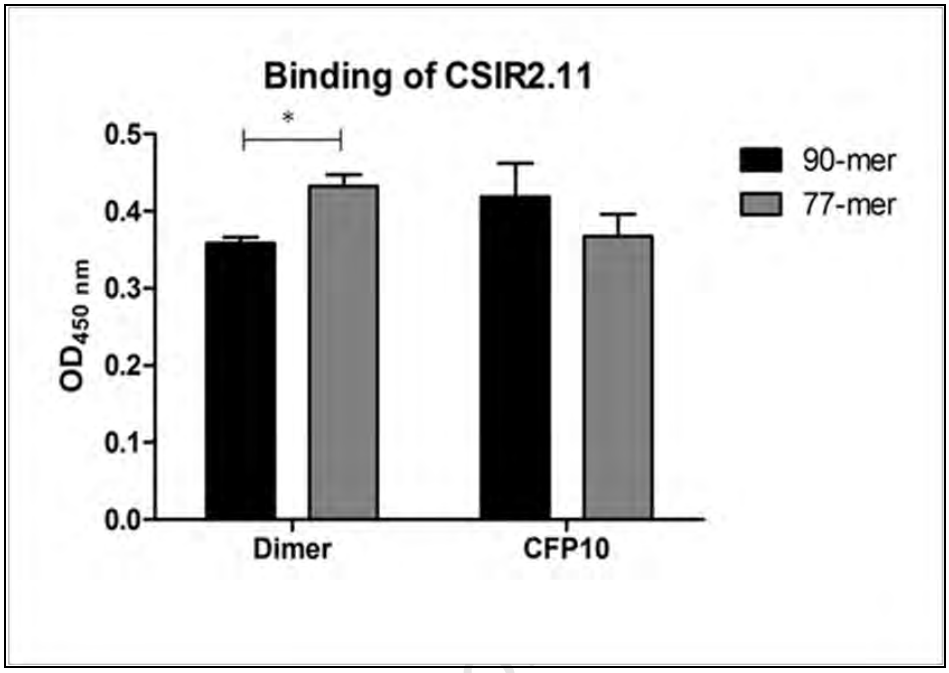


Figure A-2: Binding of CSIR 2.11 to the ESAT-6/CFP-10 dimer and CFP-10 monomer assessed by ELONA. Relative binding of the biotin-labelled aptamers to the target proteins was measured by streptavidin-HRP-catalysed reactions, detected with TMB substrate, and read as absorbance values at 450nm.The bars represent relative binding affinity of the full length, 90-mer (black), and the truncated version, 77-mer (grey), of the CSIR 2.11 aptamer. The 77-mer bound to the target proteins with affinity values comparable to those of the original 90-mer aptamer, and showed significantly higher affinity (* P-value <0.05) for the ESAT-6/CFP-10 dimer than did the 90-mer.

APPENDIX A

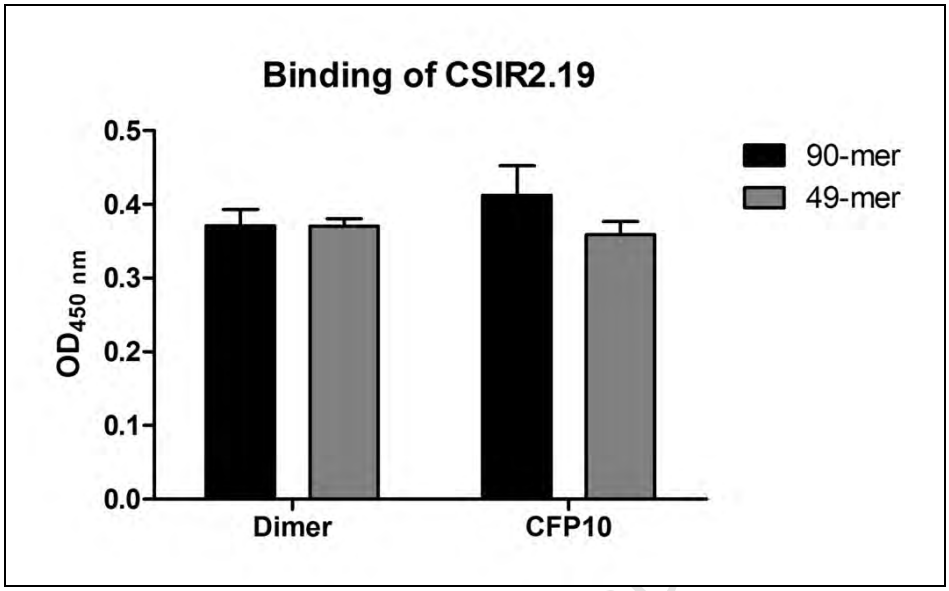


Figure A-3: Binding of CSIR 2.19 to the ESAT-6/CFP-10 dimer and CFP-10 monomer assessed by ELONA. Relative binding of the biotin-labelled aptamers to the target proteins was measured by streptavidin-HRP-catalysed reactions, detected with TMB substrate, and read as absorbance values at 450nm.The bars represent relative binding affinity of the full length, 90-mer (black), and the truncated version,49-mer (grey), of the CSIR 2.11 aptamer. The 49-mer bound to the target proteins with affinity values comparable to those of the original 90-mer aptamer for both with no statistical significance (* P-value <0.05) for both ESAT-6/CFP-10 dimer and the CFP-10 monomer.

BIBLIOGRAPHY

ALVES, C. & CUNHA, C. 2012. Electrophoretic Mobility Shift Assay: Analyzing Protein - Nucleic Acid Interactions, Gel Electrophoresis - Advanced Techniques. *In:* MAGDELDIN, D. S. (ed.) *InTech.*

ANCEL, L. W. & FONTANA, W. 2000. Plasticity, evolvability, and modularity in RNA. *J Exp Zool,* 288**,** 242-83.

ANDERSEN, P., MUNK, M. E., POLLOCK, J. M. & DOHERTY, T. M. 2000. Specific immune-based diagnosis of tuberculosis. *Lancet,* 356**,** 1099-104.

ANTONENKA, U., HOFMANN-THIEL, S., TURAEV, L., ESENALIEVA, A., ABDULLOEVA, M., SAHALCHYK, E., ALNOUR, T. & HOFFMANN, H. 2013. Comparison of Xpert MTB/RIF with ProbeTec ET DTB and COBAS TaqMan MTB for direct detection of M. tuberculosis complex in respiratory specimens. *BMC Infect Dis,* 13**,** 280.

AVCI-ADALI, M., PAUL, A., WILHELM, N., ZIEMER, G. & WENDEL, H. 2009. Upgrading SELEX technology by using lambda exonuclease digestion for single-stranded DNA generation. *Molecules,* 15**,** 1-11.

AVIRAM, I., VELTMAN, I., CHURKIN, A. & BARASH, D. 2012. Efficient procedures for the numerical simulation of mid-size RNA kinetics. *Algorithms Mol Biol,* 7**,** 24.

BAI, H., WANG, R., HARGIS, B., LU, H. & LI, Y. 2012. A SPR aptasensor for detection of avian influenza virus H5N1. *Sensors (Basel),* 12**,** 12506-18.

BIACORE 1997. *BIAevaluation 3.0 Software Handbook*.

BING, T., YANG, X., MEI, H., CAO, Z. & SHANGGUAN, D. 2010. Conservative secondary structure motif of streptavidin-binding aptamers generated by different laboratories. *Bioorg Med Chem,* 18**,** 1798-805.

BOEHME, C. C., NABETA, P., HILLEMANN, D., NICOL, M. P., SHENAI, S., KRAPP, F., ALLEN, J., TAHIRLI, R., BLAKEMORE, R., RUSTOMJEE, R., MILOVIC, A., JONES, M., O'BRIEN, S. M., PERSING, D. H., RUESCH-GERDES, S., GOTUZZO, E., RODRIGUES, C., ALLAND, D. & PERKINS, M. D. 2010. Rapid molecular detection of tuberculosis and rifampin resistance. *N Engl J Med,* 363**,** 1005-15.

BORNHORST, J. & FALKE, J. 2000. [16] Purification of Proteins Using Polyhistidine Affinity Tags. *Methods in Enzymology,* 326 245-254.

BURATOWSKI, S. & CHODOSH, L. A. 2001. Mobility shift DNA-binding assay using gel electrophoresis. *Curr Protoc Mol Biol,* Chapter 12**,** Unit 12 2.

BIBLIOGRAPHY

CHAMIE, G., LUETKEMEYER, A., CHARLEBOIS, E. & HAVLIR, D. V. 2010. Tuberculosis as part of the natural history of HIV infection in developing countries. *Clin Infect Dis,* 50 Suppl 3**,** S245-54.

CHANG, K., LU, W., WANG, J., ZHANG, K., JIA, S., LI, F., DENG, S. & CHEN, M. 2012. Rapid and effective diagnosis of tuberculosis and rifampicin resistance with Xpert MTB/RIF assay: a meta-analysis. *J Infect,* 64**,** 580-8.

CHANG, Y., DONOVAN, M. & TAN, W. 2013. Using Aptamers for Cancer Biomarker Discovery. *Journal of Nucleic Acids*,Volume 2013 Article ID 817350, 7 pages.

CHAUDHARY, V. K., KULSHRESHTA, A., GUPTA, G., VERMA, N., KUMARI, S., SHARMA, S. K., GUPTA, A. & TYAGI, A. K. 2005. Expression and purification of recombinant 38-kDa and Mtb81 antigens of Mycobacterium tuberculosis for application in serodiagnosis. *Protein Expr Purif,* 40**,** 169-76.

CHIU, T., HUANG, C. 2009. Aptamer-functionalized Nano-Biosensors. *Sensors,* 9**,** 10356-10388.

CHORLEY, B. N., WANG, X., CAMPBELL, M. R., PITTMAN, G. S., NOUREDDINE, M. A. & BELL, D. A. 2008. Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies. *Mutat Res,* 659**,** 147-57.

COWPERTHWAITE, M. C. & ELLINGTON, A. D. 2008. Bioinformatic analysis of the contribution of primer sequences to aptamer structures. *J Mol Evol,* 67**,** 95-102.

DAVIS, J. L., HUANG, L., WORODRIA, W., MASUR, H., CATTAMANCHI, A., HUBER, C., MILLER, C., CONVILLE, P. S., MURRAY, P. & KOVACS, J. A. 2011. Nucleic acid amplification tests for diagnosis of smear-negative TB in a high HIV-prevalence setting: a prospective cohort study. *PLoS One,* 6**,** e16321.

DE JONGE, M. I., PEHAU-ARNAUDET, G., FRETZ, M. M., ROMAIN, F., BOTTAI, D., BRODIN, P., HONORE, N., MARCHAL, G., JISKOOT, W., ENGLAND, P., COLE, S. T. & BROSCH, R. 2007. ESAT-6 from Mycobacterium tuberculosis dissociates from its putative chaperone CFP-10 under acidic conditions and exhibits membrane-lysing activity. *J Bacteriol,* 189**,** 6028-34.

DE MORAES VAN-LUME, D. S., CABRAL M.M.L., REGO J.C., BALBINO V., SAAD M.H., SCHINDLER H.C., ABATH F.G.C., MONTENEGRO S.M.L. 2010. Immunological diagnosis of tuberculosis based on recombinant antigens esat-6 and cfp-10 in children from an endemic area in northeast-brazil. *Scandinavian Journal of Immunology,* doi: 10.111/j.1365-3083.2010.02459.x.

DEMARSE, N. A., PONNUSAMY, S., SPICER, E. K., APOHAN, E., BAATZ, J. E., OGRETMEN, B. & DAVIES, C. 2009. Direct binding of glyceraldehyde 3-

phosphate dehydrogenase to telomeric DNA protects telomeres against chemotherapy-induced rapid degradation. *J Mol Biol,* 394**,** 789-803.

DEY, A. K., GRIFFITHS, C., LEA, S. M. & JAMES, W. 2005. Structural characterization of an anti-gp120 RNA aptamer that neutralizes R5 strains of HIV-1. *Rna,* 11**,** 873-84.

DHEDA, K., VAN ZYL SMIT, R., BADRI, M. & PAI, M. 2009. T-cell interferon-gamma release assays for the rapid immunodiagnosis of tuberculosis: clinical utility in high-burden vs. low-burden settings. *Curr Opin Pulm Med,* 15**,** 188-200.

DINIC, L., IDIGBE, O. E., MELONI, S., RAWIZZA, H., AKANDE, P., EISEN, G., ONWUJEKWE, D., AGBAJI, O., ANI, A. & KANKI, P. J. 2013. Sputum smear concentration may misidentify acid-fast bacilli as Mycobacterium tuberculosis in HIV-infected patients. *J Acquir Immune Defic Syndr,* 63**,** 168-77.

DORMAN, S. E. 2010. New diagnostic tests for tuberculosis: bench, bedside, and beyond. *Clin Infect Dis,* 50 Suppl 3**,** S173-7.

DROLET, D. W., MOON-MCDERMOTT, L. & ROMIG, T. S. 1996. An enzyme-linked oligonucleotide assay. *Nat Biotechnol,* 14**,** 1021-5.

DUPONT, D. M., MADSEN, J. B., HARTMANN, R. K., TAVITIAN, B., DUCONGE, F., KJEMS, J. & ANDREASEN, P. A. 2010. Serum-stable RNA aptamers to urokinase-type plasminogen activator blocking receptor binding. *RNA,* 16**,** 2360-9.

ELLINGTON, A. & SZOSTAK, J. 1990a. In vitro selection of RNA molecules that bind specific ligands. *Nature,* 346**,** 818-22.

ELLINGTON, A. D. & SZOSTAK, J. W. 1990b. In vitro selection of RNA molecules that bind specific ligands. *Nature,* 346**,** 818-22.

ERNST, J. D., TREVEJO-NUNEZ, G. & BANAIEE, N. 2007. Genomics and the evolution, pathogenesis, and diagnosis of tuberculosis. *J Clin Invest,* 117**,** 1738-45.

FAMULOK, M., HARTIG, J. S. & MAYER, G. 2007. Functional aptamers and aptazymes in biotechnology, diagnostics, and therapy. *Chem Rev,* 107**,** 3715-43.

FANG, X., SEN, A., VICENS, M. & TAN, W. 2003. Synthetic DNA aptamers to detect protein molecular variants in a high-throughput fluorescence quenching assay. *Chembiochem,* 4**,** 829-34.

FENG, T. T., SHOU, C. M., SHEN, L., QIAN, Y., WU, Z. G., FAN, J., ZHANG, Y. Z., TANG, Y. W., WU, N. P., LU, H. Z. & YAO, H. P. 2011. Novel monoclonal

antibodies to ESAT-6 and CFP-10 antigens for ELISA-based diagnosis of pleural tuberculosis. *Int J Tuberc Lung Dis,* 15**,** 804-10.

FLAMM, C., FONTANA, W., HOFACKER, I. & SCHUSTER, P. 2000. RNA folding at elementary step resolution. *RNA* 6**,** 325-338.

FLAMM , C., HOFACKER , I. & STADLER, P. 2004. Computational chemistry with RNA secondary structures.

FLAMM, C., HOFACKER, I., STADLER, P. & WOLFINGER, M. 2002. Barrier Trees of Degenerate Landscapes. *Z. Phys. Chem.,* 216**,** 155.

FUKUDA, T., MATSUMURA, T., ATO, M., HAMASAKI, M., NISHIUCHI, Y., MURAKAMI, Y., MAEDA, Y., YOSHIMORI, T., MATSUMOTO, S., KOBAYASHI, K., KINOSHITA, T. & MORITA, Y. S. 2013. Critical roles for lipomannan and lipoarabinomannan in cell wall integrity of mycobacteria and pathogenesis of tuberculosis. *MBio,* 4**,** e00472-12.

GAO, Y., WOLF, L. K. & GEORGIADIS, R. M. 2006. Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison. *Nucleic Acids Res,* 34**,** 3370-7.

GOLD, L., JANJIC, N., JARVIS, T., SCHNEIDER, D., WALKER, J. J., WILCOX, S. K. & ZICHI, D. 2012. Aptamers and the RNA world, past and present. *Cold Spring Harb Perspect Biol,* 4.

GOPINATH, S. 2010. Biosensing applications of surface plasmon resonance-based Biacore technology. *Sensors and Actuators B: Chemical*, 150**,** 722-733.

GREEN, L., C., B. & N., J. 2001. Aptamers as reagents for high-throughput screening. *Biotechniques,* 30**,** 1094-6, 1098.

GRUBER, A. R., BERNHART, S. H., HOFACKER, I. L. & WASHIETL, S. 2008. Strategies for measuring evolutionary conservation of RNA secondary structures. *BMC Bioinformatics,* 9**,** 122.

HALL, B., HESSELBERTH, J. R. & ELLINGTON, A. D. 2007. Computational selection of nucleic acid biosensors via a slip structure model. *Biosens Bioelectron,* 22**,** 1939-47.

HEINE, C., SCHEUERMANN, G., FLAMM, C., HOFACKER, I. L. & STADLER, P. F. 2006. Visualization of barrier tree sequences. *IEEE Trans Vis Comput Graph,* 12**,** 781-8.

HELLMAN, L. M. & FRIED, M. G. 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc,* 2**,** 1849-61.

HERMANN, T. & PATEL, D. J. 2000. Adaptive recognition by nucleic acid aptamers. *Science,* 287**,** 820-5.

BIBLIOGRAPHY

HIGUCHI, R. G. & OCHMAN, H. 1989. Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. *Nucleic Acids Res,* 17**,** 5865.

HOFACKER, I. L., FLAMM, C., HEINE, C., WOLFINGER, M. T., SCHEUERMANN, G. & STADLER, P. F. 2010. BarMap: RNA folding on dynamic energy landscapes. *Rna,* 16**,** 1308-16.

IKEBUKURO, K., KIYOHARA, C. & SODE, K. 2004. Novel electrochemical sensor system for protein using the aptamers in sandwich manner. *Biosensors and Bioelectronics.* 20, 2168-2172.

JAYASENA, S. 1999. Aptamers: An Emerging Class of Molecules That Rival Antibodies in Diagnostics. *Clinical Chemistry,* 45**,** 1628-1650.

JHAVERI, S., RAJENDRAN, M. & ELLINGTON, A. D. 2000. In vitro selection of signaling aptamers. *Nat Biotechnol,* 18**,** 1293-7.

JONES, S., DALEY, D. T., LUSCOMBE, N. M., BERMAN, H. M. & THORNTON, J. M. 2001. Protein-RNA interactions: a structural analysis. *Nucleic Acids Res,* 29**,** 943-54.

JOSS, L., MORTON, T. A., DOYLE, M. L. & MYSZKA, D. G. 1998. Interpreting kinetic rate constants from optical biosensor data recorded on a decaying surface. *Anal Biochem,* 261**,** 203-10.

KANWAR, J. R., ROY, K. & KANWAR, R. K. 2011. Chimeric aptamers in cancer cell-targeted drug delivery. *Crit Rev Biochem Mol Biol,* 46**,** 459-77.

KAUFMANN, S. H. & PARIDA, S. K. 2008. Tuberculosis in Africa: learning from pathogenesis for biomarker identification. *Cell Host Microbe,* 4**,** 219-28.

KHAN, H., MISSAILIDIS, S. 2008. Aptamers in oncology: a diagnostic perspective. *Gene Therapy and Molecular Biology,* 12.

KHATI, M., SCHUMAN, M., IBRAHIM, J., SATTENTAU, Q., GORDON, S. & JAMES, W. 2003. Neutralization of infectivity of diverse R5 clinical isolates of human immunodeficiency virus type 1 by gp120-binding 2'F-RNA aptamers. *J Virol,* 77**,** 12692-8.

KIM, K. S., LEE, H. S., YANG, J. A., JO, M. H. & HAHN, S. K. 2009a. The fabrication, characterization and application of aptamer-functionalized Si-nanowire FET biosensors. *Nanotechnology,* 20**,** 235501.

KIM, M., UM, H., BANG, S., LEE, S., OH, S., HAN, J., KIM, K., MIN, J. & YH., K. 2009b. Arsenic removal from Vietnamese groundwater using the arsenic-binding DNA aptamer. *Environ Sci Technol.,* 43**,** 9335-40.

BIBLIOGRAPHY

KIM, Y., LIU, C. & TAN, W. 2009c. Aptamers generated by Cell SELEX for biomarker discovery. *Biomark Med,* 3**,** 193-202.

KIMPLE, A. J., MULLER, R. E., SIDEROVSKI, D. P. & WILLARD, F. S. 2010. A capture coupling method for the covalent immobilization of hexahistidine tagged proteins for surface plasmon resonance. *Methods Mol Biol,* 627**,** 91-100.

KNECHEL, N. A. 2009. Tuberculosis: pathophysiology, clinical features, and diagnosis. *Crit Care Nurse,* 29**,** 34-43; quiz 44.

KULSHRESTHA, A., GUPTA, A., VERMA, N., SHARMA, S. K., TYAGI, A. K. & CHAUDHARY, V. K. 2005. Expression and purification of recombinant antigens of Mycobacterium tuberculosis for application in serodiagnosis. *Protein Expr Purif,* 44**,** 75-85.

LABARRE, P., HAWKINS, K. R., GERLACH, J., WILMOTH, J., BEDDOE, A., SINGLETON, J., BOYLE, D. & WEIGL, B. 2011. A simple, inexpensive device for nucleic acid amplification without electricity-toward instrument-free molecular diagnostics in low-resource settings. *PLoS One,* 6**,** e19738.

LANIEL, M. A., BELIVEAU, A. & GUERIN, S. L. 2001. Electrophoretic mobility shift assays for the analysis of DNA-protein interactions. *Methods Mol Biol,* 148**,** 13-30.

LEE, J. H., YIGIT, M. V., MAZUMDAR, D. & LU, Y. 2010. Molecular diagnostic and drug delivery agents based on aptamer-nanomaterial conjugates. *Adv Drug Deliv Rev,* 62**,** 592-605.

LEE, S. J., YOUN, B. S., PARK, J. W., NIAZI, J. H., KIM, Y. S. & GU, M. B. 2008. ssDNA aptamer-based surface plasmon resonance biosensor for the detection of retinol binding protein 4 for the early diagnosis of type 2 diabetes. *Anal Chem,* 80**,** 2867-73.

LEONTIS, N. B., LESCOUTE, A. & WESTHOF, E. 2006. The building blocks and motifs of RNA architecture. *Curr Opin Struct Biol,* 16**,** 279-87.

LEWIS, K. N., LIAO, R., GUINN, K. M., HICKEY, M. J., SMITH, S., BEHR, M. A. & SHERMAN, D. R. 2003. Deletion of RD1 from Mycobacterium tuberculosis mimics bacille Calmette-Guerin attenuation. *J Infect Dis,* 187**,** 117-23.

LI, J. J., FANG, X. & TAN, W. 2002. Molecular aptamer beacons for real-time protein recognition. *Biochem Biophys Res Commun,* 292**,** 31-40.

LI, Y., LEE, H. J. & CORN, R. M. 2006. Fabrication and characterization of RNA aptamer microarrays for the study of protein-aptamer interactions with SPR imaging. *Nucleic Acids Res,* 34**,** 6416-24.

BIBLIOGRAPHY

LISS, M., PETERSEN, B., WOLF, H. & PROHASKA, E. 2002. An aptamer-based quartz crystal protein biosensor. *Anal Chem,* 74**,** 4488-95.

LIU, G., MAO, X., PHILLIPS, J. A., XU, H., TAN, W. & ZENG, L. 2009. Aptamer-nanoparticle strip biosensor for sensitive detection of cancer cells. *Anal Chem,* 81**,** 10013-8.

LIU, J. & LU, Y. 2006. Preparation of aptamer-linked gold nanoparticle purple aggregates for colorimetric sensing of analytes. *Nat Protoc,* 1**,** 246-52.

LIU, J., MAZUMDAR, D. & LU, Y. 2006. A simple and sensitive "dipstick" test in serum based on lateral flow separation of aptamer-linked nanostructures. *Angew Chem Int Ed Engl,* 45**,** 7955-9.

LORENZ, R., BERNHART, S. H., HONER ZU SIEDERDISSEN, C., TAFER, H., FLAMM, C., STADLER, P. F. & HOFACKER, I. L. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol,* 6**,** 26.

LUZI, E., MINUNNI, M., TOMBELLI, S. & MASCINI, M. 2003. New trends in affinity sensing: aptamers for ligand binding. *TrAC Trends in Analytical Chemistry,* 22**,** 810-818.

MALLIKARATCHY, P. R., RUGGIERO, A., GARDNER, J. R., KURYAVYI, V., MAGUIRE, W. F., HEANEY, M. L., MCDEVITT, M. R., PATEL, D. J. & SCHEINBERG, D. A. 2011. A multivalent DNA aptamer specific for the B-cell receptor on human lymphoma and leukemia. *Nucleic Acids Res,* 39**,** 2458-69.

MARIMUTHU, C., TANG, T. H., TOMINAGA, J., TAN, S. C. & GOPINATH, S. C. 2012. Single-stranded DNA (ssDNA) production in DNA aptamer generation. *Analyst,* 137**,** 1307-15.

MARRO, M. L., DANIELS, D. A., MCNAMEE, A., ANDREW, D. P., CHAPMAN, T. D., JIANG, M. S., WU, Z., SMITH, J. L., PATEL, K. K. & GEARING, K. L. 2005. Identification of potent and selective RNA antagonists of the IFN-gamma-inducible CXCL10 chemokine. *Biochemistry,* 44**,** 8449-60.

MATHEBULA, N. S., PILLAY, J., TOSCHI, G., VERSCHOOR, J. A. & OZOEMENA, K. I. 2009. Recognition of anti-mycolic acid antibody at self-assembled mycolic acid antigens on a gold electrode: a potential impedimetric immunosensing platform for active tuberculosis. *Chem Commun (Camb)***,** 3345-7.

MATHEWS, D. H. & TURNER, D. H. 2006. Prediction of RNA secondary structure by free energy minimization. *Curr Opin Struct Biol,* 16**,** 270-8.

MAYER, K. H. & DUKES HAMILTON, C. 2010. Synergistic pandemics: confronting the global HIV and tuberculosis epidemics. *Clin Infect Dis,* 50 Suppl 3**,** S67-70.

BIBLIOGRAPHY

MICHAEL T. WOLFINGER, M., ANDREAS SVRCEK-SEILER, W., FLAMM, C., HOFACKER, I. & STADLER, P. 2004. Exact Folding Dynamics of RNA Secondary Structures. *J. Phys. A: Math. Gen,* 37

MOORE, M. D., COOKSON, J., COVENTRY, V. K., SPROAT, B., RABE, L., CRANSTON, R. D., MCGOWAN, I. & JAMES, W. 2011. Protection of HIV neutralizing aptamers against rectal and vaginal nucleases: implications for RNA-based therapeutics. *J Biol Chem,* 286**,** 2526-35.

MURPHY, M. B., FULLER, S. T., RICHARDSON, P. M. & DOYLE, S. A. 2003. An improved method for the in vitro evolution of aptamers and applications in protein detection and purification. *Nucleic Acids Res,* 31**,** e110.

NEVES, M. A., REINSTEIN, O. & JOHNSON, P. E. 2010. Defining a stem length-dependent binding mechanism for the cocaine-binding aptamer. A combined NMR and calorimetry study. *Biochemistry,* 49**,** 8478-87.

NG, N., SHIMA, D., CALIAS, P., CUNNINGHAM, E. J., GUYER, D. & ADAMIS, A. 2006. Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *NATURE REVIEWS | DRUG DISCOVERY,*5, 123-132.

NGUYEN, T., HILTON, J. & LIN, Q. 2009. Emerging applications of aptamers to micro- and nanoscale biosensing. *Microfluid Nanofluid,*2009**,** 347-362

NICOL, M. P. & ZAR, H. J. 2011. New specimens and laboratory diagnostics for childhood pulmonary TB: progress and prospects. *Paediatr Respir Rev,* 12**,** 16-21.

NONAKA, Y., SODE, K. & IKEBUKURO, K. 2010. Screening and improvement of an anti-VEGF DNA aptamer. *Molecules,* 15**,** 215-25.

PAI, M. & LING, D. I. 2008. Rapid diagnosis of extrapulmonary tuberculosis using nucleic acid amplification tests: what is the evidence? *Future Microbiol,* 3**,** 1-4.

PAI, M., ZWERLING, A. & MENZIES, D. 2008. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. *Ann Intern Med,* 149**,** 177-84.

PARIDA, S. K. & KAUFMANN, S. H. 2010. The quest for biomarkers in tuberculosis. *Drug Discov Today,* 15**,** 148-57.

PATZEL, V. 2004. In silico selection of functional RNA molecules. *Curr Opin Drug Discov Devel,* 7**,** 360-9.

PENDERGRAST, P. S., MARSH, H. N., GRATE, D., HEALY, J. M. & STANTON, M. 2005. Nucleic acid aptamers for target validation and therapeutic applications. *J Biomol Tech,* 16**,** 224-34.

BIBLIOGRAPHY

POTTY, A. S. R., KOURENTZI K, FANG H, JACKSON G.W, XING ZHANG, LEGGE G.B, WILLSON R.C 2008. Biophysical Characterization of DNA Aptamer Interactions with Vascular Endothelial Growth Factor

POTYRAILO, R. A., CONRAD, R. C., ELLINGTON, A. D. & HIEFTJE, G. M. 1998. Adapting selected nucleic acid ligands (aptamers) to biosensors. *Anal Chem,* 70**,** 3419-25.

PROSKE, D., BLANK, M., BUHMANN, R. & RESCH, A. 2005. Aptamers—basic research, drug development, and clinical applications. *Applied Microbiology and Biotechnology,* 69**,** 367-374.

PYM, A. S., BRODIN, P., BROSCH, R., HUERRE, M. & COLE, S. T. 2002. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti. *Mol Microbiol,* 46**,** 709-17.

PYM, A. S., BRODIN, P., MAJLESSI, L., BROSCH, R., DEMANGEL, C., WILLIAMS, A., GRIFFITHS, K. E., MARCHAL, G., LECLERC, C. & COLE, S. T. 2003. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nat Med,* 9**,** 533-9.

RENSHAW, P. S., PANAGIOTIDOU, P., WHELAN, A., GORDON, S. V., HEWINSON, R. G., WILLIAMSON, R. A. & CARR, M. D. 2002. Conclusive evidence that the major T-cell antigens of the Mycobacterium tuberculosis complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6*CFP-10 complex. Implications for pathogenesis and virulence. *J Biol Chem,* 277**,** 21598-603.

ROBERTSON, D. L. & JOYCE, G. F. 1990. Selection in vitro of an RNA enzyme that specifically cleaves single- stranded DNA. *Nature,* 344**,** 467-8.

ROCKEY, W., HERNANDEZ, F., HUANG, S., CAO, S., HOWELL, C., THOMAS, G., LIU, X., LAPTEVA, N., SPENCER, D., MCNAMARA, J., ZOU, X., CHEN, S. & GIANGRANDE, P. 2011. Rational truncation of an RNA aptamer to prostate-specific membrane antigen using computational structural modeling. *Nucleic Acid Ther. ,* 21**,** 299-314.

ROTHERHAM, L. S., MASERUMULE, C., DHEDA, K., THERON, J. & KHATI, M. 2012. Selection and application of ssDNA aptamers to detect active TB from sputum samples. *PLoS One,* 7**,** e46862.

RUIGROK, V. J., VAN DUIJN, E., BARENDREGT, A., DYER, K., TAINER, J. A., STOLTENBURG, R., STREHLITZ, B., LEVISSON, M., SMIDT, H. & VAN DER OOST, J. 2012. Kinetic and stoichiometric characterisation of streptavidin-binding aptamers. *Chembiochem,* 13**,** 829-36.

BIBLIOGRAPHY

RUSCONI, C. P., ROBERTS, J. D., PITOC, G. A., NIMJEE, S. M., WHITE, R. R., QUICK, G., JR., SCARDINO, E., FAY, W. P. & SULLENGER, B. A. 2004. Antidote-mediated control of an anticoagulant aptamer in vivo. *Nat Biotechnol,* 22**,** 1423-8.

SANDMAN, K., KRZYCKI, J., DOBRINSKI, B., LURZ, R., REEVE, J. & (1990). 1990. HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon Methanothermus fervidus, is most closely related to histone. *Proceedings of the National Academy of Sciences of the United States of America,* 87**,** 5788-5791.

SANTALUCIA, J., JR. & HICKS, D. 2004. The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct,* 33**,** 415-40.

SAVORY, N., ABE, K., SODE, K. & IKEBUKURO, K. 2010. Selection of DNA aptamer against prostate specific antigen using a genetic algorithm and application to sensing. *Biosens Bioelectron.*

SCHUSTER, P. 2009. Genotypes and Phenotypes in the Evolution of Molecules. *European Review,* 17**,** 281-319

SHANGGUAN, D. & TANG Z., M. P., XIAO Z., TAN W. 2007. Optimization and Modifications of Aptamers Selected from Live Cancer Cell Lines. *ChemBioChem,* 8**,** 603 - 606.

SHIGDAR, S., WARD, A., DE, A., YANG, C., WEI, M. & DUAN, W. 2011. Clinical applications of aptamers and nucleic acid therapeutics in haematological malignancies. *Br J Haematol. ,* 155**,** 3-13.

SIMMONS, S. C., MCKENZIE, E. A., HARRIS, L. K., APLIN, J. D., BRENCHLEY, P. E., VELASCO-GARCIA, M. N. & MISSAILIDIS, S. 2012. Development of novel single-stranded nucleic acid aptamers against the pro-angiogenic and metastatic enzyme heparanase (HPSE1). *PLoS One,* 7**,** e37938.

SIVASHANMUGAM, A., MURRAY, V., CUI, C., ZHANG, Y., WANG, J. & LI, Q. 2009. Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Sci. ,* 18**,** 936-948.

SLUTSKY, M. & MIRNY, L. A. 2004. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys J,* 87**,** 4021-35.

SMITH, I. 2003. Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. *Clin Microbiol Rev,* 16**,** 463-96.

SOLOVIC, I., JONSSON, J., KORZENIEWSKA-KOSELA, M., CHIOTAN, D. I., PACE-ASCIAK, A., SLUMP, E., RUMETSHOFER, R., ABUBAKAR, I., KOS, S., SVETINA-SORLI, P., HAAS, W., BAUER, T., SANDGREN, A. & VAN

BIBLIOGRAPHY

DER WERF, M. J. 2013. Challenges in diagnosing extrapulmonary tuberculosis in the European Union, 2011. *Euro Surveill,* 18.

STOJANOVIC, M. N., DE PRADA, P. & LANDRY, D. W. 2001. Aptamer-based folding fluorescent sensor for cocaine. *J Am Chem Soc,* 123**,** 4928-31.

STOLTENBURG, R., NIKOLAUS, N. & STREHLITZ, B. 2012. Capture-SELEX: Selection of DNA Aptamers for Aminoglycoside Antibiotics. *J Anal Methods Chem,* 2012**,** 415697.

STOLTENBURG, R., REINEMANN, C. & STREHLITZ, B. 2007. SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol Eng,* 24 381-403.

STORLA, D. G., YIMER, S. & BJUNE, G. A. 2008. A systematic review of delay in the diagnosis and treatment of tuberculosis. *BMC Public Health,* 8**,** 15.

STREHLITZ, B., NIKOLAUS, N. & STOLTENBURG, R. 2008. Protein Detection with Aptamer Biosensors. *Sensors* 8**,** 4296-4307.

TOMBELLI, S., MINUNNI, M., LUZI, E. & MASCINI, M. 2005. Aptamer-based biosensors for the detection of HIV-1 Tat protein. *Bioelectrochemistry,* 67**,** 135-41.

TUERK, C. & GOLD, L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science,* 249**,** 505-10.

WANG, Y., LI, D., REN, W., LIU, Z., DONG, S. & WANG, E. 2008. Ultrasensitive colorimetric detection of protein by aptamer-Au nanoparticles conjugates based on a dot-blot assay. *Chem Commun (Camb)***,** 2520-2.

WARD, W. & SWIATEK, G. 2009. Protein Purification.*Current Analytical Chemistry,* 5 85-105(21)

WASHIETL, S., WILL, S., HENDRIX, D. A., GOFF, L. A., RINN, J. L., BERGER, B. & KELLIS, M. 2012. Computational analysis of noncoding RNAs. *Wiley Interdiscip Rev RNA,* 3**,** 759-78.

WHO 2012. Global Tuberculosis Report 2012. Geneva: World Health Organization: World Health Organization.

WIN, M., KLEIN, J. & SMOLKE, C. 2006. Codeine-binding RNA aptamers and rapid determination of their binding constants using a direct coupling surface plasmon resonance assay. *Nucleic Acids Res. ,* 34**,** 5670-5682.

WOLFINGER, M., ANDREAS SVRCEK-SEILER, W., FLAMM, C., HOFACKER, I. & STADLER, P. 2004. Exact Folding Dynamics of RNA Secondary Structures. *J. Phys. A: Math. Gen,* 37

BIBLIOGRAPHY

WUCHTY, S., FONTANA, W., HOFACKER, I. & SCHUSTER, P. 1998. Complete Suboptimal Folding of RNA and the Stability of Secondary Structure. *Biopolymers,* Vol 49.

WUCHTY S, F. W., HOFACKER I.L, SCHUSTER P 1998. Complete Suboptimal Folding of RNA and the Stability of Secondary Structure. *Biopolymers,* Vol 49.

XIA, F., ZUO, X., YANG, R., XIAO, Y., KANG, D., VALLEE-BELISLE, A., GONG, X., YUEN, J. D., HSU, B. B., HEEGER, A. J. & PLAXCO, K. W. 2010. Colorimetric detection of DNA, small molecules, proteins, and ions using unmodified gold nanoparticles and conjugated polyelectrolytes. *Proc Natl Acad Sci U S A,* 107**,** 10837-41.

XIE, S. & WALTON, S. P. 2010. Development of a dual-aptamer-based multiplex protein biosensor. *Biosens Bioelectron,* 25**,** 2663-8.

XU, H., MAO, X., ZENG, Q., WANG, S., KAWDE, A. N. & LIU, G. 2009. Aptamer-functionalized gold nanoparticles as probes in a dry-reagent strip biosensor for protein analysis. *Anal Chem,* 81**,** 669-75.

ZEKA, A. N., TASBAKAN, S. & CAVUSOGLU, C. 2011. Evaluation of the GeneXpert MTB/RIF assay for rapid diagnosis of tuberculosis and detection of rifampin resistance in pulmonary and extrapulmonary specimens. *J Clin Microbiol,* 49**,** 4138-41.

ZHOU, J., SOONTORNWORAJIT, B., SNIPES M.P., WANG., Y 2010. Structural Prediction and Binding Analysis of Hybridised Aptamers. *Journal of Molecular Recognition,* Volume 24**,** 119-126.

ZHU, D., ZHOU, X. & XING, D. 2010. A new kind of aptamer-based immunomagnetic electrochemiluminescence assay for quantitative detection of protein. *Biosens Bioelectron,* 26**,** 285-8.

ZUKER, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res,* 31**,** 3406-15.

ZUKER, M. & SANKOFF, D. 1984. RNA SECONDARY STRUCTURES AND THEIR PREDICTION. *Bulletin of Mathematical Biology,* 46**,** 591-621.